

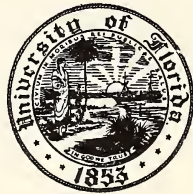
THOMAS
SILVERMAN
and
SARAH
SILVERMAN

**Cognitive
Techniques
in Marketing
Analysis**

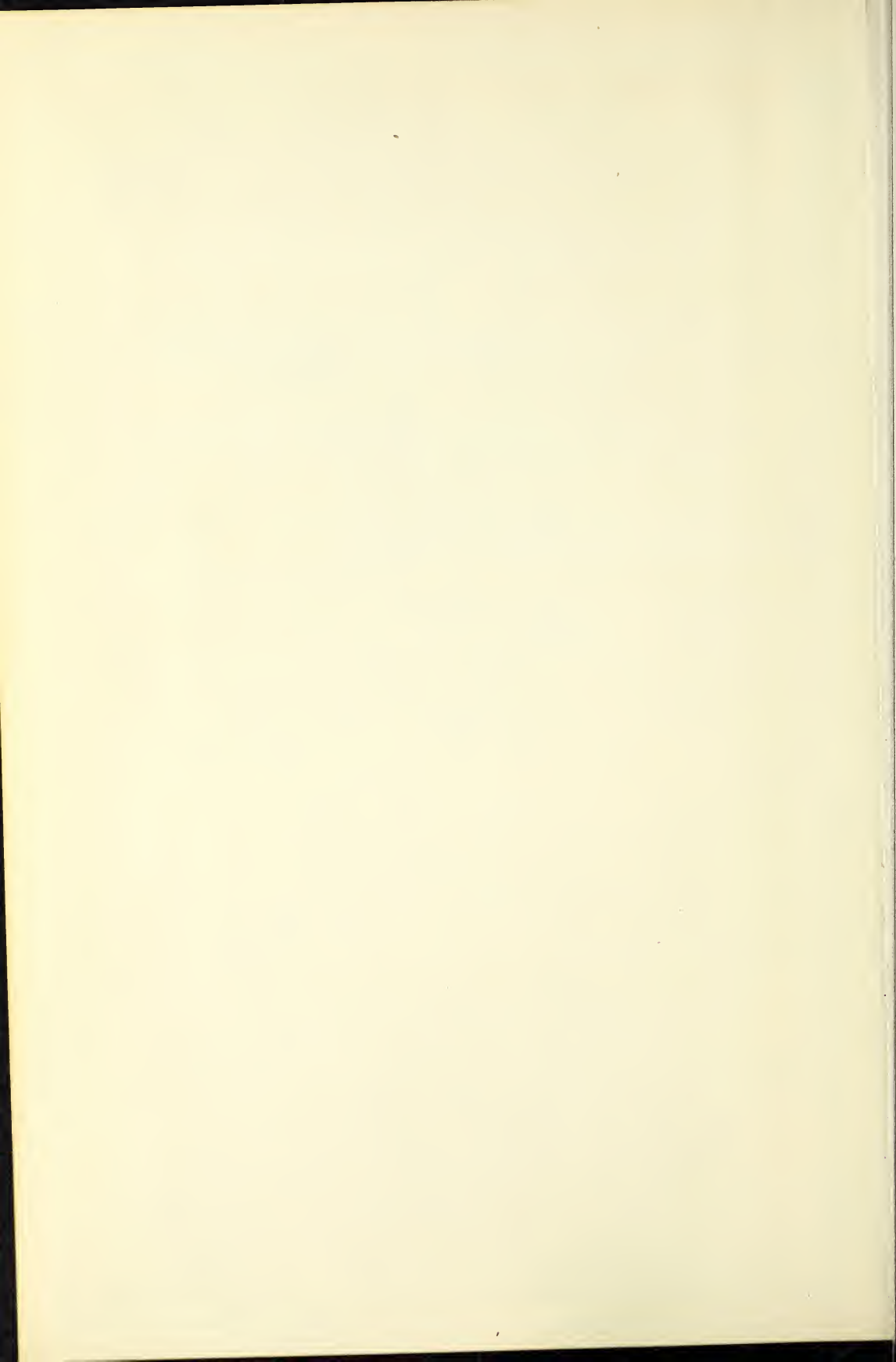
Text and Readings

From *Journal
of
Consumer
Research*
and *Journal of
Marketing*

UNIVERSITY
OF FLORIDA
LIBRARIES







**QUANTITATIVE
TECHNIQUES
IN MARKETING
ANALYSIS**

.....
Text and Readings

IRWIN SERIES IN QUANTITATIVE ANALYSIS FOR BUSINESS

Consulting Editor

ROBERT B. FETTER *Yale University*

FETTER & DALLECK *Decision Models for Inventory Management*

BOWMAN & FETTER *Analysis for Production Management* Re-
vised Edition

BIERMAN, FOURAKER, & JAEDICKE *Quantitative Analysis for Busi-
ness Decisions*

MARTIN *Electronic Data Processing*

BASS, BUZZELL, GREENE, LAZER, PESSEMIER, SHAWVER, SHUCHMAN,
THEODORE, & WILSON *Mathematical Models and Methods
in Marketing*

MORRIS *Analysis for Materials Handling Management*

FRANK, KUEHN, & MASSY *Quantitative Techniques in Marketing
Analysis: Text and Readings*

QUANTITATIVE TECHNIQUES IN MARKETING ANALYSIS

TEXT AND READINGS

Ronald E. Frank, Ph.D. *Assistant Professor of Business Administration, Harvard University*

Alfred A. Kuehn, Ph.D. *Associate Professor of Industrial Administration, Carnegie Institute of Technology*

William F. Massy, Ph.D. *Assistant Professor of Marketing, Stanford University*

1962 Richard D. Irwin, Inc. Homewood, Ill.

© 1962 BY RICHARD D. IRWIN, INC.

ALL RIGHTS RESERVED. THIS BOOK OR ANY PART THEREOF MAY NOT BE
REPRODUCED WITHOUT THE WRITTEN PERMISSION OF THE PUBLISHER

First Printing, September, 1962

Library of Congress Catalogue Card No. 62-20666

PRINTED IN THE UNITED STATES OF AMERICA

Bus. ad.

B + D

1-2-63

.....to Richard M. Cyert



Preface

THE MATERIAL CONTAINED IN THIS volume can be looked upon as reflecting a feeling of the authors as to one of the principal trends that is occurring in the field of marketing (and will continue to occur over the long run). In our opinion the analysis of complex marketing problems will gradually move in the direction of increased reliance upon a rather extensive array of quantitative techniques. Eventually these techniques may become as familiar to the marketing analyst as are break-even analysis and sampling theory now.

We have grouped the techniques into three categories: experimentation, statistical analysis, and simulation. While it is quite likely that a given research effort is apt to draw upon the types of techniques in more than one of these categories, it seems nonetheless pedagogically useful to make this separation, as it serves to emphasize the differences in the underlying logic of the various approaches.

In addition to this trend toward quantitative techniques, the articles selected for inclusion reflect three other trends, which are at different stages of development. The interdisciplinary character of our selections reflects the fact that an increased fraction of new entrants into the field of marketing is coming from backgrounds with a strong disciplinary bent (for example, economics and econometrics, statistics and mathematics, and psychology and sociology). Scholars in other disciplines are also showing an increasing interest in marketing problems. Last, but not least, our selection reflects what we believe to be an increasing international exchange with respect to research bearing on marketing problems (for example, see the articles by Belson, Ehrenberg, Gridgeman, Harper, and Jurén).

In our papers we have chosen to emphasize the underlying logic involved in creation and implementation of each of the respective techniques. We do not wish to get involved in the labyrinth of testing techniques so closely associated with classical statistics—however valuable they may be in certain contexts. It is our hope that the materials will be used in classes with widely varying prerequisites with respect to statistics. Those groups with a relatively modest training will probably want to concentrate to a greater extent upon the adequacy of each of the respective models as reflectors of the underlying marketing problem. The more advanced the background of the students the more emphasis might be placed on the manipulation as well as the construction of the respective models.

The primary bases for selecting the articles were: (1) the extent to which a particular technique appeared to be relevant to the solution of some set of marketing problems, (2) the extent to which it is currently being applied, and (3) the extent to which we feel that it will receive increased application in the future. Where good illustrations of analysis could be found within the field of marketing, they were chosen in preference to material from other disciplines. In situations where we felt that a particular technique was of importance, but where adequate examples were not available within the usual marketing literature, we searched for an example from some related discipline. For example, we wanted several articles illustrating the use of simulation as a technique for handling rather complex problems. We were unable to find a suitable set of expositions in the usual marketing literature. Rather than delete consideration of this tool, we chose instead to draw upon articles dealing with subject matter usually treated by economists (Orcutt's paper) or by psychologists and political scientists (Abelson and Pool).

We are indebted to Dean Richard M. Cyert of the Graduate School of Business Administration, Carnegie Institute of Technology for the original idea which led to this book. In addition to the authors whose works we are reprinting, we would like to thank Professors Henry J. Claycamp of M.I.T., Ralph L. Day of the University of Texas, John B. Matthews of the Harvard Business School, and Mr. Robert G. Shaw of the Carnegie Institute of Technology for their careful review of portions of the manuscript and extremely helpful suggestions. Sally Orcutt and Alice Morse deserve special recognition for their contributions to the preparation of the manuscript. Finally, we owe a vote of thanks to our wives, Iris, Heidi, and June, for their patience and assistance during a most trying period.

RONALD E. FRANK
ALFRED A. KUEHN
WILLIAM F. MASSY

Cambridge, Massachusetts
June 27, 1962

Contents

	PAGE
Introduction, WROE ALDERSON	xi
PART I. THE ROLE OF RESEARCH	1
The Role of Research in Marketing Management—HARRY V. ROBERTS, <i>Journal of Marketing</i>	3
PART II. FUNDAMENTALS	19
Research Design in Marketing Analysis—RONALD E. FRANK	21
— Statistical Analysis of Relations between Variables—WIL- LIAM F. MASSY	56
— Complex Interactive Models—ALFRED A. KUEHN	106
PART III. READINGS ON EXPERIMENTATION	125
A. LABORATORY EXPERIMENTATION	127
Basic Research Report on Consumer Behavior—ALDERSON AND SESSIONS	129
A Tasting Experiment—NORMAN THEODORE GRIDGEMAN, <i>Applied Statistics</i>	146
An Experimental Method for Estimating Demand—EDGAR A. PESSEMIER, <i>Journal of Business</i>	153
Why Television Commercials Succeed—HORACE S. SCHWER- IN, <i>Motivation and Market Behavior</i>	166
B. FIELD EXPERIMENTATION	177
The Measurement of the Effect of a New Packaging Ma- terial upon Preference and Sales—SEYMOUR BANKS, <i>Jour- nal of Business</i>	179
A Switch-Over Experimental Design to Measure Advertis- ing Effect—R. J. JESSEN, <i>Journal of Advertising Research</i>	190
Sales Effects of Two Campaign Themes—PETER L. HEN- DERSON, JAMES F. HIND, and SIDNEY E. BROWN, <i>Journal of Advertising Research</i>	204
A Technique for Studying the Effects of a Television Broadcast—WILLIAM A. BELSON, <i>Applied Statistics</i>	220
PART IV. READINGS ON STATISTICAL ANALYSIS	229
A. CROSS-CLASSIFICATION ANALYSIS	231
Research Analysis—A. J. WOOD RESEARCH CORPORATION, <i>Wood Chips</i>	233
Social Processes in Physicians' Adoption of a New Drug— JAMES COLEMAN, HERBERT MENZEL and ELIHU KATZ, <i>Journal of Chronic Diseases</i>	239
B. REGRESSION AND DISCRIMINANT ANALYSIS	265
Some Correlates of Coffee and Cleanser Brand Shares— SEYMOUR BANKS, <i>Journal of Advertising Research</i>	269

	PAGE
Long-Term Trends in Food Consumption: A Multi-Country Study—L. JURÉEN, <i>Econometrica</i>	281
Psychological and Objective Factors in the Prediction of Brand Choice; Ford versus Chevrolet—FRANKLIN B. EVANS, <i>Journal of Business</i>	303
C. PROBABILITY MODELS FOR BRAND LOYALTY	341
Early Prediction of Market Success for New Grocery Products—LOUIS A. FOURT and JOSEPH W. WOODLOCK, <i>Journal of Marketing</i>	345
The Pattern of Consumer Purchases—A. S. C. EHRENBERG, <i>Applied Statistics</i>	355
Brand Choice as a Probability Process—RONALD E. FRANK, <i>Journal of Business</i>	372
Consumer Brand Choice—A Learning Process?—ALFRED A. KUEHN	390
D. MULTIPLE FACTOR ANALYSIS	405
Factor Analysis as a Technique for Examining Complex Data on Foodstuffs—ROLAND HARPER, <i>Applied Statistics</i>	409
A Multiple Factor Analysis of Advertising Readership—DIK WARREN TWEDT, <i>Journal of Applied Psychology</i>	427
Television Ownership in 1950: Results of a Factor Analytic Study—WILLIAM F. MASSY	440
PART V. READINGS ON SIMULATION	461
A. SIMULATION	463
The Simulmatics Project—ITHIEL DE SOLA POOL and ROBERT ABELSON, <i>Public Opinion Quarterly</i>	465
A Stochastic Microanalytic Model of a Socioeconomic System—GUY H. ORCUTT, MARTIN GREENBERGER, JOHN KORBEL, and ALICE M. RIVLIN, <i>Microanalysis of Socioeconomic Systems; A Simulation Study</i>	481
A Model of Retail Ordering and Pricing by a Department Store—R. M. CYERT, J. G. MARCH, and C. G. MOORE	502
A Heuristic Program for Locating Warehouses—ALFRED A. KUEHN and MICHAEL J. HAMBURGER	523
INDEX	549

Introduction

by Wroe Alderson

THE MARKETING MANAGER PURSUES two broad objectives on behalf of his company—namely, market expansion and marketing efficiency. The company goals can be assumed to be some combination of growth and profits. Both are dependent on the success of the firm's marketing function.

A company grows by selling more products or by selling to more customers. Market analysis undertakes to identify potential customers and to find the means for reaching them through advertising and selling. Market expansion would not lead to profits if it cost too much to gain new customers. The total cost of market expansion might be defined to include research and development for creating new products or improving old ones. The marketing manager, however, is more directly responsible for marketing efficiency. He uses tools such as advertising and selling in his effort to expand the market, but he must make efficient use of these sales dollars if increased sales are to mean increased profits.

Research analysis aids the pursuit of marketing efficiency in various ways. Among others, it helps to identify the best distribution channels and methods for a given product and the most effective combination or marketing mix to achieve the desired results.

The use of systematic marketing analysis in the United States is scarcely more than 50 years old. Over that entire period the market analyst has struggled with two fundamental research problems. One is to get the facts and the other is to analyze the facts in a way that is helpful to executive judgment. During most of the 50-year history of research, it concentrated on the first task of obtaining the marketing facts.

The emphasis on fact finding as compared to analysis was doubtless due primarily to two conditions. One was the very serious lack of knowledge on the part of the marketing executive of the actual conditions surrounding the ultimate sale of his product. In those days when a product was offered through the channels of trade, all that a sponsor ever really knew was that the product either succeeded or failed. He did not know what kind of people bought it, in what type of store they made the purchase, or at what point repeat purchases began to compare favorably with trial purchases. He did not know what appeals attracted the consumer to the product in the first place, what they liked or disliked about

it once they had tried it, or how much the market would have been increased or decreased by offering the product at a different price. Similarly there were great blind spots with respect to the kinds of stores available, the best methods of obtaining cooperation from these stores, or the number and types of dealers which should be combined so as to get the most economical distribution for the product. In short, the sales manager of that day almost completely lacked all of the types of basic information which seem essential for market planning today.

It is interesting to note that the first approach to systematic marketing research took place in two areas of great uncertainty concerning the ultimate consumer-buyer. One of these areas was the advertising business and, specifically, The Curtis Publishing Company. The advertiser of a consumer product was by definition marketing at a distance to people whom he would never meet and about whom he knew very little. His only sales contacts were with the wholesaler and retailer, but his advertising went to consumers who would normally buy the product in retail stores.

It was under these conditions that Charles Coolidge Parlin began the search for marketing facts as a service to Curtis advertisers. He anticipated many of the massive fact-finding enterprises of a later day. His Department Store Study, for example, was a sort of census of large retailers which anticipated the first official Census of Distribution by many years. His study of city markets was the first attempt to produce buying power maps for major metropolitan centers.

Another source of interest in marketing facts was the problems of agricultural marketing. The growers of specified crops, distributing them throughout the United States, were another example of marketing at a distance.

The second commercial research organization to be established, following that of Curtis in 1911, was that of Swift & Company in 1917. Further impetus from the agricultural side came in 1921 when Congress authorized what was to be known as the Joint Commission of Agricultural Inquiry. This study examined the reasons for the marketing spread between the farmer and the consumer but incidentally showed for the first time the intricacies of channels of distribution for many important products. Other major fact-finding efforts include the first Census of Distribution in 1929 and the chain store inquiry by the Federal Trade Commission, extending over some years during the same period.

It has been said that marketing analysis sprang from a paucity of information as one of its major causes. Another reason for concentrating on fact finding rather than analysis was a general skepticism as to whether the more refined analytical techniques available had any application in the business world.

There is a tendency today to speak of marketing research and analysis as applied economics, but that is certainly not the way it developed. Gen-

eral economists at the time were not interested in the details of marketing processes or marketing channels. As a matter of fact, by assuming in their equations that information was costless and universally available, they swept aside most of what was important or interesting in marketing. Here and there economists appeared with an interest in concrete business problems, and they sometimes made striking demonstrations of the value of economic analysis in such areas as marketing management. Generally, however, marketing and economics were developing along very different lines in the universities, and market analysts in business developed their own separate body of techniques.

The analysis of marketing data was largely conventional and sometimes quite elementary. Some analysts struck out in new directions such as the development of distribution cost analysis, first in the U.S. Department of Commerce, with extensions and refinements to business and the universities. Most of the techniques employed represented nothing more than a transfer of production cost accounting to the marketing field. The crudity of these methods and their alleged failure to deal with marginal costs scandalizes some economists to this day. The fact that these methods usually worked and had demonstrated time after time their capacity to contribute to marketing efficiency merely seemed to compound the outrage. These earlier developments with respect to fact finding and analysis might be called the conventional or pre-formal phase of the development.

The main effort in marketing was certainly quantitative in the sense that it dealt with numbers rather than with logical structure alone, as in the economic theory of the period. The term "quantitative techniques" today has come to imply a more advanced body of methodology. Nevertheless the really basic problems of fact finding and analysis and the relation of one to the other were as much in evidence then as at the present time. It will be worthwhile to comment on this earlier debate before turning to the situation which confronts the analyst in applying more formal quantitative techniques today.

In general it was recognized that fact finding and analysis go hand in hand. There were differences of emphasis which can be dramatized in terms of the extreme positions. At one pole were those who said, "Get the facts and the facts will speak for themselves." At the other pole were those who said, "A problem well stated is half-solved. In fact, it may yield to logical analysis with little or no resort to fact finding." At one time the market researcher and the general economist could be said to stand at these opposite poles but later the same debate was to continue among the marketing practitioners themselves. At one period Alfred Politz and Ernest Dichter were the most vocal champions of positions approaching these extremes. One stressed measurement of consumer behavior and attitudes, the other asserted that consumer motivation was the really basic marketing problem to be attacked by depth interviews. One relied on

large probability samples and asserted that unless data were statistically valid they were misleading and of no value to management. The other said that no amount of descriptive data on overt behavior or surface attitudes could be of any value if the real problem was motivations which were hidden even from the subjects themselves.

The most extreme positions of all are those which are attributed to either side by their opponents. Thus the measurement people are charged with paying attention only to those variables which can readily be counted or measured and leaving out the only variables which really matter. The retort is that motivation specialists have greatly exaggerated the importance of certain intangibles and have supported their position by a handful of striking cases which may not be at all representative of consumers as a whole. Motivation research is taken here only as an example. The broader contrast is between quantifiable variables and intangibles which are intuitively felt to be controlling.

Marketing analysis is now moving into a new stage of methodological awareness which brings it closer to what is recognized as science in other fields. The key concept in this new development is that of the explicit analytical model. A model can be described by specifying a set of variables and the relations which exist among these variables. Market analysts who have tried to solve problems have always behaved as if they were using some type of model even though they never bothered to say what the model was. In the past these models have sometimes been extremely simple, and that is one reason why it may not have seemed necessary to make them explicit.

Many years ago a marketing publication by the federal government presented native white population as the primary measure of markets, county by county. The underlying model was extremely simple indeed, and the author of the publication might have been shocked to have it made explicit. The model said in effect that white people born in the United States buy goods while Negroes and persons of foreign extraction do not. Marketing has made a great deal of use of such classificatory models, but many are more refined and make use of more data than the one just described.

Explicit models now used in marketing analysis more typically embrace continuous variables and a structure of relationships showing how one variable affects another. This type of model is usually judged by its power to predict. For example, given certain inputs such as advertising and selling into the marketing system represented by the model, it should be possible to predict certain outcomes such as sales and profits. Where a model had predictive value, however, it must include all the significant variables, and the functional relationships among them must be correctly specified.

It should be noted that explicit models can still take account of simple classifications such as native white population, but they do not end there.

The contrast between quantitative and qualitative information is not nearly so sharp as some people on either side would like to make it. If there is a certain class of consumers differing qualitatively from other consumers, it may be possible to specify the way in which they are different, and it is then at least theoretically feasible to count the number of people who belong in this class. Thus the number of left-handed men can be brought into a sophisticated marketing model just as readily as the number of men who are precisely 6 feet tall. This would be impossible only if the analyst merely asserted that left-handedness would affect the use of a given product but refused to count or estimate the number of left-handed men.

Situations of precisely this character have occurred in marketing in the relatively recent past. A rather famous motivation research study concluded that some consumers associated animal fats with a sense of sin and that this would affect their judgment of certain shortening compounds. This fact could have been built into a marketing model except that the report failed to provide two significant figures. One was the number or percentage of consumers in the market who might be expected to have such an idea. The other was the relative strength of this idea for those who held it, in comparison to other factors of motivation affecting their purchases of shortening compounds.

Those who proposed to rely on intuitive or qualitative distinctions for solving marketing problems often refer to their opponents as nose counters. The measurement specialists might retort that they were not interested in counting noses as such, but simply contended that it was essential to count or measure any variable which had a significant bearing on the problem in order to obtain data that would be useful in solving it. The limitation of the measurement viewpoint is that some of its proponents have stopped short at this point rather than using their numerical data in an explicit quantitative model.

It would be a mistake to assume that quantitative models contribute only to the analytical rather than the data side of problem solving. Quantitative models are not the purely logical models of economic theory in which the only results to be obtained are inferences drawn from the structure of the model itself. The very term "quantitative model" suggests the combination of empirical data with the logical structure of the system which the model represents. Instead of perpetuating the argument between the fact finders and the logicians, the new emphasis on explicit quantitative models holds out the hope that these separate contributions can now be successfully integrated. In fact, the quantitative models which are now coming into use should lead to improvements in both the logic and the data of marketing analysis.

The coming of electronic computers and the possibilities for simulating complex systems enables the analyst to cope with a far more intricate logic than was ever possible before. The logic of mathematical economics

remained very general for two reasons. First, the economic theorist was primarily interested in very broad and abstract relationships. Secondly, when the same methods were applied to problems in the micro-economic field it was not generally feasible to derive quantitative solutions from a set of equations. The reason was that the number of variables and the number of relationships that had to be included in a formulation of the marketing situation were usually so numerous as to place precise solutions beyond the range of available mathematical techniques. Computer simulation jumps clear over this barrier to technical progress. The number of variables and quantitative relationships which the computer can handle is significantly greater than the number we can identify and measure in a marketing situation.

Similarly, the use of explicit models seems bound to lead to a great advance in fact finding and measurement. The use of an explicit model provides the researcher with a much more definite criterion as to the data to be collected. It allows him to test for the degree of precision in measuring a given variable which is required for his analysis. In the case of some variables which have never been measured before, he can start out with estimates as to how the variable will behave within a given range and then take steps to make these values progressively more accurate.

One major way in which explicit models will influence both data gathering and analysis is in placing greater stress on functional relationships rather than merely on the measurement of the significant variables. The researcher will be less concerned with the measurement obtained under given circumstances than in the way the variable may be expected to move as the conditions change. There was less incentive in the past for trying to define these functional relationships since they presently become too numerous to handle by the available techniques.

One other aspect of the debate over methodology in marketing is worthy of comment. This is the view that marketing is by nature an applied behavioral science and that the new quantitative approach tends to substitute mechanism for the insights of behavioral science. It must be admitted that this is occasionally true, but it certainly is not necessarily true. The analyst who regards a marketing system as a wholly deterministic mechanism may get the wrong answers because he has neglected the importance of people at every level in the system. An operating system is also an organization which is organized around the goals of the total operation, but it is organized by and for the objectives and incentives of the people who make it up.

The prudent analyst will avoid the extremes of looking only at the mechanistic aspects which can most readily be put into a model or abandoning explicit models entirely because they mirror the real world imperfectly. For the time being his best strategy is to put into the model every variable which he thinks can have a real effect on the outcome, whether he is able to measure it or is obliged to guess at it. If there are parts

of organizational behavior left over that he can find no way of putting into the model he should be able to deal with them more effectively than if he were working with no model at all.

First he might look at the outcome predicted by the model as it stands. Secondly, he could undertake to make some judgments as to how the results may differ from those predicted because of the factors left out. Working along this line he may eventually find ways of quantifying additional intangibles and thus round out the basic structure of the model. Far from interfering with the application of behavioral science to marketing, the use of an explicit model enables us to say things about marketing and about human behavior generally that we were never able to say so well before. The most important thing to say about an aspect of behavior is how it relates to other aspects of behavior. The use of an explicit model forces the analyst in the direction of estimating such functional relationships and then gathering data for confirmation or refinement.

Neither marketing nor behavioral science can advance purely in terms of existential statements that such and such behavior can be found among consumers or among marketers. What interests us in marketing is how these aspects of behavior work together to produce a result. If we regard ourselves as cultural anthropologists we may delight in certain patterns of behavior simply because they are curious or extreme. As marketing men we would like to draw on cultural anthropology as well as the other behavioral sciences, but our role and purpose is quite different. It is our function to contribute to the basic objectives of marketing organizations which have already been mentioned—namely, marketing expansion and marketing efficiency. Certainly we must specify these results in quantitative terms in order to measure our own performance. Similarly we must find ways of quantifying all the significant variables and relating them in explicit models if our predictions and our recommendations are to contribute to these marketing objectives.

University of Pennsylvania
July, 1962



THE ROLE OF RESEARCH

PART ONE



The Role of Research in Marketing Management*

HARRY V. ROBERTS†

EXPANSION OF MARKETING RESEARCH REFLECTS A GROWING BELIEF THAT the methods of science are useful in solving the problems of business management. Effective application of marketing research, however, is neither easy nor automatic; and some have even contended that on balance actual applications have been more ineffective than effective.¹ This paper represents an attempt to formulate a framework for analysis of conditions under which marketing research can be expected to be effective and to make incidental suggestions for increasing effectiveness. This framework is founded on *a priori* reasoning and impressionistic evidence; its presentation here is drastically condensed; but it may nonetheless be suggestive both to those who apply research and to those who are intrigued by research for its own sake.

DECISION MAKING

Marketing management can be viewed simply as the continuing attempt to recognize and solve specific marketing problems. A *problem* exists when an objective is desired and there is uncertainty as to how it can best be achieved. A *decision* is the selection of some course of action (or inaction) to attain the objective. There are three imperatives in the process of decision making: (1) possible actions must be recognized; (2) the results of different actions must be predicted; and (3) the order of preference of these predicted results must be assessed. Research is potentially useful for at least the first two of these.

* Reprinted from the *Journal of Marketing*, national quarterly publication of the American Marketing Association, Vol. XXII, No. 1 (July, 1957), pp. 21-32.

† University of Chicago.

¹ For example, see John E. Jeuck, "Marketing Research—Milestone or Millstone?" *The Journal of Marketing*, April, 1953, pp. 381-87.

RESEARCH AND INTUITION

"Research" can be contrasted with "intuition" (or "judgment" or "common sense") in the decision-making process.

Research is any relatively systematic, formal, conscious procedure for evolving and testing hypotheses about reality or, in more modern terms, for making decisions. The words "systematic," "formal," and "conscious" differentiate "research" from "intuition." The distinction between research and intuition is not a sharp one, especially since intuition is an essential ingredient of good research, but the major differences of emphasis are these:

1. While both research and intuition are ultimately oriented toward predictions, intuition is oriented toward narrow and immediate predictions rather than general hypotheses fruitful of many specific predictions.
2. Intuition is learned, if at all, by experience and demonstration rather than by formal study.
3. Intuition is seldom subject to logical scrutiny or formal empirical testing.
4. Research uses such technical tools as mathematics, logic, experimental methods, and statistical inference.

Since research is intuition plus science, it is easy to assume that research can improve upon intuition in the social sciences and business management as well. There is, however, reason to be skeptical of this assumption. A distinguished economist has said,

We seem to be forced to the conclusion, not that prediction and control are impossible in the field of human phenomena, but that the formal methods of science are of very limited application. Common sense does predict and control, and can be trained to predict and control better; but that does not prove that science can predict and control better than common sense. And it seems very doubtful whether in the majority of social problems the application of logical methods and canons will give as good results as the informal, intuitive process of judgment which, when refined and developed, becomes art.²

In order to formulate useful generalizations about the role of research in marketing, two closely related issues must be considered: (1) potential contributions of research to marketing problems and (2) identification of characteristics of marketing problems that make these problems more or less accessible to attack by research.

POTENTIAL CONTRIBUTIONS OF RESEARCH

There are four major kinds of contributions that research can make toward decision making in marketing: (1) systematic description and classification of marketing "facts" as exemplified by the censuses of population and distribution or the data of the Audit Bureau of Circulation; (2) substantive hypotheses that can be used to make predictions; (3) logical and mathematical tools for classifying relevant variables, exploring the

² Frank H. Knight, *The Ethics of Competition and Other Essays* (London: George Allen and Unwin, Ltd., 1936), pp. 132-33. See also pp. 116-17 and 119.

logical relationships among hypotheses, and deriving the predictions implied by hypotheses; and (4) the inferential tools of statistics.

Systematic Description and Classification of "Facts"

It is easy to forget and hard to evaluate precisely the contribution of a wealth of descriptive materials to marketing management in America, and indeed these materials are often not even thought of as "marketing research" because of their descriptive rather than analytical orientation. Yet, a little reflection on the pervasive use in marketing, for example, of census materials suggests their indispensable role.

Substantive Hypotheses

In physics the body of substantive hypotheses is large, relatively well articulated by logical connections between individual hypotheses, and well verified by actual testing of the accuracy of predictions derived from hypotheses. In short, one may speak meaningfully of "physical theory." By contrast, the so-called "social" or "behavioral" sciences, which have the greatest potential relevance to marketing, *consist largely of apparently unrelated empirical observations with relatively little theory to articulate these observations*. These disciplines rarely contain hypotheses of direct usefulness in solving the specific problems which arise in marketing. Moreover, most current marketing research is *ad hoc* empirical work with little substantive carry-over from one study to another. The so-called "principles" in textbooks rarely yield concrete predictions as to the effect of proposed marketing actions. For example, "principles" of copywriting and advertising format are often listed, but advertising men frequently disagree on the choice of the most effective advertisement.³

The main theoretical instrument for potential guidance in marketing is the axiom of rational behavior, which has been applied mainly in economics.⁴ Supply and demand theory based on the rationality postulates has some relevance for marketing problems. For example, even without precise quantitative estimates of elasticities of demand for automobiles, this theory will predict the effect of setting the retail list price of cars below the competitive price on (a) the distribution of income between dealers and manufacturers, (b) the rate of production of automobiles, and (c) the rate of sale to consumers of "optional" accessories.⁵

Even though there are few predictive hypotheses in the social sciences,

³ Most textbook principles are really proverbs; for each "principle" there is an equally plausible "principle" stating the opposite. See Herbert A. Simon, *Administrative Behavior* (New York: The Macmillan Co., 1948), pp. 20 ff.

⁴ Kenneth J. Arrow, "Mathematical Models in the Social Sciences," *Cowles Commission Papers, New Series, No. 48* (Chicago: Cowles Commission for Research in Economics, 1952), p. 137.

⁵ Milton Friedman, "Notes on Lectures in Price Theory" (unpublished manuscript based on notes prepared by David I. Fand and Warren J. Gustus, The University of Chicago Press, 1951), pp. 15-16.

these sciences may often be useful in marketing by suggesting what to look for as opposed to predicting what will be found. Thus, a psychologist might be able to suggest possible advertising appeals that would not occur to a copywriter who lacked a formal background in psychology, even though psychological theory probably could not predict the most effective from a given list of appeals.

Logical and Mathematical Tools

Logical and mathematical tools are essentially languages for the analysis of problems. The concepts of supply and demand schedules, for example, would facilitate the analysis of economic—and some marketing—problems even if they had no predictive value at all. Pure formal logic may also be helpful. For example, there is one advertising “principle” that the advertising message should reach as wide an audience as possible and another that frequent repetitions of the advertising message are desirable. Logic suggests immediately that either “principle” by itself is inadequate and that the real problem is to find some optimal combination of wide coverage and intensive repetition. Similar examples are easy to find; in all of them, logic serves to formulate the real issues and avoid futile talking.

In recent years many new mathematical tools have been made available for possible use in making marketing decisions. For example, various mathematical models have been proposed for problems of strategy, programming, learning, inventory decisions, mass and small group communication, economic forecasting, queuing, capital budgeting, diversification of risk, and so on. Actual applications have undoubtedly lagged behind advance publicity, but enough has been done to suggest exciting developments to come.

Statistical Inference

Statistical inference is potentially useful in marketing as a tool for securing and interpreting empirical observations so that a “rational” decision can be made in the face of uncertainty. In *experimental* applications, managerial actions are actually tried out with the aim of discovering the responses to these actions. All other applications are *nonexperimental* or “*observational*.”

Evidence cast up by experience is abundant and frequently as conclusive as that from contrived experiments. . . . But such evidence is far more difficult to interpret. It is frequently complex and always indirect and incomplete . . . its interpretation generally requires subtle analysis and involved chains of reasoning, which seldom carry real conviction. . . .⁶

⁶ Milton Friedman, *Essays in Positive Economics* (Chicago: The University of Chicago Press, 1953), pp. 10–11.

Experimental methods, by contrast, yield evidence free of such difficulties of interpretation so long as: (1) The underlying conditions of the past persist into the future. (2) The experiment is run sufficiently long that responses to experimental stimuli will have time to manifest themselves. (3) The population being studied can be broken down into smaller units (families, stores, sales territories, etc.) for which the experimental stimuli can be measured and for which responses to the stimuli are not "contagious." (4) The experimenter is able to apply or withhold, as he chooses, experimental stimuli from any particular unit of the population he is studying. (5) Neither the stimulus nor the response is changed by the fact that an experiment is being conducted. (6) The sample size is large enough to measure important responses to experimental stimuli against the background of uncontrolled sources of variation.

The key to modern statistical design of experiments is withholding experimental stimuli *at random*. To the extent that randomization and the other conditions above are met, the responses actually observed will reflect the "true" effects of the stimuli plus random or chance variation. Statistical procedures then need cope only with the interpretation of chance variation. Observational methods must cope both with chance variation and possible systematic error, and systematic error eludes rigorous statistical treatment. For example, one can compare purchases of an advertiser's product between readers and nonreaders of a recent advertisement and try to make statistical allowance for the fact that readers are potentially different from nonreaders in many respects other than the simple fact of seeing or not seeing the advertisement; but after all such allowances, there is no guarantee that systematic differences do not persist. For example, a statistician might make statistical adjustments for factors such as economic status, education, sex, age, etc., that might distinguish readers from nonreaders of the advertisement and then find that the adjusted purchase rate of the product is still higher among readers than nonreaders. *Even if* the statistical adjustments have succeeded in making the two groups comparable in respect to *all* variables associated with the tendency to respond to the advertisement's message—and there is never any guarantee in observational studies that this has in fact been achieved—it still may be true that those people who buy the product are more likely to notice and read the advertisement or at least to remember, or say that they remember, having read it.

These fundamental limitations of observational techniques notwithstanding, marketing research has already drawn heavily on them and will do so increasingly in the future. The development of sound methods of sampling of human populations is recent, and advances are still being made. Much of current marketing research has been made possible by this development of sampling theory and practice. Similarly, the development of psychological scaling and factor analysis—which may be viewed as statistical tools—has had a noticeable influence on current practice.

Moreover, there are many potentially valuable statistical tools that have not yet been widely applied in marketing research, and others will undoubtedly be developed.

CHARACTERISTICS OF MARKETING PROBLEMS THAT INFLUENCE ATTAINMENT OF AN ECONOMICAL SOLUTION BY RESEARCH

Every marketing problem is different, but certain criteria determine the compatibility of particular problems and methods of research. In the simplest type of problem situation, the proposed actions are given. The response to the actions must be predicted and a decision made in the light of the desired objectives. General economic conditions, actions of competitors, and other actions of the firm are assumed to continue unchanged. Later these assumptions will be relaxed. In each of the criteria listed below, the phrase "other things being equal" should be considered implicit.

The more rapid the response to marketing actions, the easier the problem for research. Often the quality characteristics of manufactured products can be measured only after a long time. Thus, the resistance to weathering of different paints could be tested by (1) actual weathering over five years or (2) accelerated testing in a laboratory. It is well known in product research that accelerated testing is capricious and often perverse.⁷ In marketing the shortcomings of accelerated testing are at least as serious but less likely to be discovered, especially for institutional and "indirect-action" advertising and promotion. By contrast, direct-action selling effort such as local and mail-order advertising—for which accelerated testing is unnecessary—has been researched widely with obvious success.⁸ In addition to the rapidity of response to direct-action selling effort, there is often a second explanation for success of research: responses are often readily *traceable*. Hence we set down:

The easier it is to trace response to marketing actions, the easier the problem for research. Responses to sales effort are frequently difficult to trace, partly because the many forms of selling effort that impinge on the ultimate consumer are often closely interrelated and many of them are designed to work indirectly. For example, an argument frequently advanced for the use of national advertising to *consumers* is that *dealers* are favorably motivated thereby. Further, consumer buying decisions may be interdependent because of "conspicuous consumption," "word-of-mouth," or "opinion leadership." There are tools for coping with the problem of traceability. Selling stimuli often can be traced to relatively small geographic areas or relatively short time periods. If the buying decisions of consumers are interdependent, the contagion of purchase may

⁷ Eugene W. Pike, "Testing Components for High-Reliability Systems" (unpublished paper given at the National Meeting on Airborne Electronics of the Institute of Radio Engineers, Dayton, Ohio, May 13, 1953).

⁸ For a typical example, see *Printers' Ink*, July 11, 1952, pp. 35 and 110.

be less important between small geographic areas than between individual consumers.

When measurement of the desired responses to marketing actions is impossible or uneconomic, the problem will be easier to the extent that satisfactory substitute responses can be found. If response to sales stimuli takes too long to manifest itself or is difficult to trace, a substitute response highly correlated with buying responses can be sought. Copy tests, consumer jury ratings, preference studies, and many techniques of questionnaire design and psychological measurement are based upon substitute responses.

In deciding on the validity of proposed substitute responses, there are three closely related approaches: (1) to verify by research that the substitute response is closely correlated with the response of ultimate interest for a class of problems similar to the one being studied, (2) to draw on general theoretical knowledge, and (3) to draw on experience and intuition.

It is hard to find validations of substitute responses in product research. In marketing, there are relatively few documented examples with details of the process of validation. Most of these examples fail to show significant differences for the sales responses but show significant differences for the corresponding substitute responses, usually preference measurements.⁹ Here is the dilemma. Should one use preference measurements that show significant differences in response to different selling stimuli but for which results are not related in any known way with sales measurements? Preference measurements may be correlated with sales responses and are perhaps more sensitive. Seymour Banks has suggested that:

. . . a preferent test can select the best of a series of proposed alternatives, but it cannot tell how much effect the use of this best alternative will have upon sales.¹⁰

But Banks has also published an excellent counterexample to this speculation.¹¹ George Brown has suggested in informal discussions that the "chain of causation" from selling stimulus to buying response might be divided into stages: opportunity for exposure to selling effort (for example, read the magazine in which an advertisement appears); actual exposure to selling effort (for example, read the actual advertisement); acqui-

⁹ See, for example: James H. Lorie and Harry V. Roberts, *Basic Methods of Marketing Research* (New York: McGraw-Hill Book Co., Inc., 1951), pp. 209-11; Seymour Banks, "The Measurement of the Effect of a New Packaging Material upon Preference and Sales," *The Journal of Business*, April, 1950, pp. 79 ff.; and G. Maxwell Ule, unpublished doctoral dissertation in progress, The University of Chicago, School of Business.

¹⁰ Seymour Banks, *ibid.*, p. 80.

¹¹ Seymour Banks, "The Prediction of Dress Purchases for a Mail-Order House," *The Journal of Business*, January, 1950, pp. 48-57.

sition of information about or confidence in the product; formation of the intention to buy; and actual purchase. Measurements can be made at any stage of this chain. One might speculate that measurements would be more valid the closer they are to actual purchase, but this is at best only a plausible speculation. In the absence of evidence, speculation is unavoidable; but the search for validation of substitute responses should continue.

To the extent that experimentation is economically feasible, the problem is easier to solve by research. Properly designed experiments are free of many of the limitations of observational studies. Businessmen generally are often better situated to conduct sound experiments than is commonly believed. Actions taken as a result of business decisions represent direct intervention of a kind that is usually impossible in the social sciences. Both "test-tube" experiments as a prelude to final marketing decision and continuing experiments thereafter are often possible and entail little added cost, especially when marketing policies can be introduced piecemeal geographically.

Randomized experimentation is impossible only for "all-or-none" decisions. Major plant construction and advertising in national media are examples. John Jeuck listed major marketing innovations made without research and concluded, ". . . the really significant marketing achievements, the big chance . . . is dependent primarily on imagination and on the skills of management."¹²

Even when randomization through space is precluded, randomization *over time* may be feasible if the full response to selling stimuli—such as special promotions undertaken intermittently at randomly chosen time periods—is rapid and if the statistical behavior of sales in the absence of experimentation is well understood.¹³ Unfortunately, the full response may not be manifest sufficiently rapidly. Thus the initial effect of premium offers may be a very substantial increase in the rate of sales followed by a decrease if the premium mainly affects timing of purchases rather than their total amount.

Finally, experimentation may be undesirable if the stimulus or its response is seriously modified by the fact that an experiment is being run.

The success of research in problems that cannot be economically attacked experimentally depends on the adequacy of observational methods. Evaluation of adequacy of observational methods is also frustrated by the absence of evidence, but plausible guides can be offered.

The more adequate the theoretical knowledge, the more adequate are observational methods based on this knowledge. When theoretical knowledge is very good, satisfactory predictions can be made on the basis of

¹² John E. Jeuck, "Marketing Research Today: A Minority Report," *School of Business Publications I*, The University of Chicago, undated, p. 10.

¹³ See R. L. Anderson, "Recent Advances in Finding Best Operating Conditions," *Journal of American Statistical Association*, Vol. XLVIII, No. 264 (1953), pp. 789-98.

existing theory without further recourse to observational study. Bridge building and other feats of engineering construction illustrate this point. More frequently, theory is simply an invaluable aid in the design and interpretation of research, whether observational or experimental. But while experimentation can reach valid results in problems for which theoretical knowledge is wholly or partly lacking, observational studies are seriously weakened when theory is sparse. There is relatively little marketing theory, and intuition must often guide the use of observational techniques. The distinction between "research" and "intuition" therefore becomes very blurred, as in certain areas of medical research in which valid experiments are difficult and theoretical knowledge is meager. In both marketing and medicine, the questions which cannot be subjected to valid experimentation are frequently resolved by the intuition of the practitioner.

Observational techniques may be more adequate to the extent that statistical allowance has been made for disturbing variables that may obscure the relationship being studied. By statistical allowance is meant crossclassification, standardized averages, multiple regression, and related techniques. These techniques are often used in observational studies to answer the question, "what would have happened if certain disturbing variables had not varied?" It seems always prudent to attempt to answer this question within the resources available for research. Except in experimental studies, of course, there is no guarantee that the answers are satisfactory within the usual allowance for chance error.¹⁴

There are a few special devices, not as widely used as they might be, that may enhance the effectiveness of observational methods in marketing research. (1) Aggregate data can be divided into components, as when sales are studied by territories or families. "Sales analysis," which seems to have lost favor by comparison with survey methods, illustrates the approach. (2) Rates of change through time can be studied instead of absolute levels. While this "before-after" approach is common in sales tests, its use in observational studies is rare, perhaps because of the current emphasis on single surveys. (3) Insightful comparisons can be made such as comparisons of product awareness of advertised and competitive brands during a special campaign or Alfred Politz's comparison of attitudes toward car acceleration with strength of accelerator spring.¹⁵ (4) Different observational methods may be used to attack a single research problem. The degree of convergence of their results may indicate the confidence to be accorded the results of any one of the approaches.

To the extent that changes in present conditions may have to be al-

¹⁴ See, for example, W. Allen Wallis and Harry V. Roberts, *Statistics: A New Approach* (Glencoe, Ill.: Free Press of Glencoe, Inc., 1956), chap. ix.

¹⁵ Alfred Politz, "Science and Truth in Marketing Research," *Harvard Business Review*, January-February, 1957, pp. 121-22.

lowed for, the problem is more difficult for research. Two cases need to be considered: (1) changes that would have occurred anyway; and (2) changes that are, at least in part, responses to the decision in the particular problem.

Changes that would have occurred anyway. The best-known examples are shifts in general business conditions or in demand conditions confronting a whole industry. General economic conditions have such relevance to marketing (and other) problems of the firm that economic and sales forecasting is of great importance. A treatment of the usefulness of research in economic forecasting is beyond our present scope.¹⁶

As to the modifications of marketing decisions that should be made in the light of predicted changes in economic conditions, economic theory is relatively useful. For example, there is considerable evidence of the difference in response of prices and outputs of durable and nondurable goods or of retail and wholesale prices to changes in economic conditions.

Changes in part influenced by the decision under consideration. The classic illustration analyzed by economic theorists is the situation of duopoly or other small-number competitive situations. The analogue in marketing is easy to find. Suppose, for example, that a company were able by experimental methods to estimate the productivity of its advertising expenditures and that a leading competitor had maintained his advertising unchanged during the period of the experiment. As a result of the experiment, the company might sharply increase its advertising expenditures, making the tacit assumption that the competitor would make no change. The competitor—observing this increase—might increase its advertising, possibly “cancelling out” the effectiveness of the first company’s added expenditure and leaving both companies in their original sales position vis-à-vis one another. The experiment would have been entirely adequate for determining what would have happened in the absence of a response by the competitor, but it could lead to a bad decision in the absence of a prediction as to what the competitor would do; how soon he would do it; and what effect, if any, his response would have on the rate of sales of the first company. Competitive responses and interactions also raise serious problems in the interpretation of observational data.

The guessing of competitive reactions may be an art highly dependent

¹⁶ See, for example: Harry V. Roberts, “A Method of Improving Sales Forecasts,” *Proceedings of Golden Triangle Conference of the American Marketing Association* (American Marketing Association, Inc., 1956), pp. 97–102; Robert Ferber, “Sales Forecasting by Correlation Methods,” *The Journal of Marketing*, January, 1954, pp. 219–32; James H. Lorie, “Operations Research and Short-Run Sales Forecasting” (paper delivered before the Second Annual Meeting of the Operations Research Society of America, May 21–22, 1954); Irving Schweiger, “The Contribution of Consumer Anticipations in Forecasting Consumer Demand,” in National Bureau of Economic Research, *Short-Term Economic Forecasting*, Studies in Income and Wealth, Vol. XVII (Princeton, N.J.: Princeton University Press, 1955).

on intuition and experience, with formal research perhaps serving chiefly to supply salient background facts. Psychology and social-psychology do not appear to offer useful predictions. Even if the competitive responses could be predicted, their effect on sales could be estimated at best by observational methods.

Finally, the task of research is easier for problems that can be studied and solved independently of other problems because indirect effects do not then have to be considered. One of the key tasks of management—and research—is the fragmentation of the complex of problems facing a company into smaller problems that can be considered relatively independently of other problems.

THE EFFECTIVENESS OF RESEARCH IN THE DISCOVERY OF POSSIBLE NEW ALTERNATIVES FOR ACTION

An essential part of problem solving is the discovery of alternative possibilities for action besides those initially considered or—in alternative terminology—the “identification of problems.” This is closely analogous to “formulation of hypotheses” where words like “accident” and “inspiration” readily come to mind. In the context of marketing, John E. Jeuck has said:

. . . It seems likely that research can do very little in a positive or creative way to lead to those new possibilities that are the essence of creative market development and exploitation. That is simply to say, I suppose, that marketing success depends more upon the imaginative and aggressive personality who may in the process of development make many errors than it does upon the careful collation of facts and the cautious investigation of alternatives that are the hallmark of research operations.

In that connection, one must be impressed with how unlikely it would have been for Sears, Roebuck ever to have made a really strong start if Richard Sears had had a clear view of what is usually uncovered in the consumer survey. One wonders how many other companies would have never been started if they had relied upon the typical consumer survey to guide them in their selection of products and policies. . . .¹⁷

To the extent that Jeuck's comments represent more than a criticism of the frequent use of more or less stereotyped methods in marketing research, they are seriously misleading. Research is probably more effective in unearthing new possibilities for action than in predicting the response to existing ones.¹⁸ Research focuses attention on possible actions that probably would not have been recognized in the absence of research. The new ideas that turn up as an unexpected by-product of research are

¹⁷ Jeuck, *op. cit.*, p. 7.

¹⁸ The impressions reported here are based on fairly extensive examination of applications of research in several areas of management. Many of these applications—about 300—were made by students in the 9th, 10th, 11th, and 12th Groups of the Executive Program of the University of Chicago who were given the assignment of applying statistical methods to problems in their own companies. See Harry V. Roberts, “Statistics in Middle Management,” *Management Science*, Vol. I, Nos. 3-4 (1955), pp. 224-32.

frequently more valuable than the objectives originally sought. The very discipline of objective recording of data forces people to look at things they would never have looked at otherwise and to question assumptions that would not otherwise have been challenged: research can be a protection against complacency and a stimulus to imagination.

Quality control illustrates a method of research aimed primarily at the discovery of new possibilities for action.¹⁹ The basic idea is that a repetitive process such as a manufacturing or marketing operation is measured at periodic intervals, often by sampling methods. Statistical procedures—"control charts"—have been devised in order to detect systematic changes of the process. The detection of such changes provides a warning that the process should be immediately investigated to see what has happened and to discover "assignable causes." By discovering the "assignable cause," management can either remove a source of trouble or retain and possibly apply more widely an improvement that otherwise might have been lost. Quality control methods, then, suggest looking for trouble and therefore opportunity at times and places where the search is likely to be rewarded.

"RESEARCH" VERSUS "INTUITION"

Many problems in business are and should be attacked without the benefit of formal methods of research—especially relatively minor and nonrecurrent problems and those that require very quick decisions. Research would then cost more than its benefits. But though intuition sometimes yields a snap decision without any reflection, it may also involve soul searching and interminable committee meetings, both of which are expensive in terms of executive time.

If the criteria for the applicability of research enumerated above are examined carefully, four seem just as applicable to intuition or informal decision making as to formal methods of research: (1) the more rapid the response to actions, (2) the more easily responses can be traced, (3) the more adequate the substitute responses that can be found, and (4) the fewer the changes in existing conditions that need to be taken into account, the easier is the problem for *either* research or intuition. The criteria that appear to discriminate between research and intuition are the economic feasibility of experimentation and, failing experimentation, the availability of adequate observational tools.

RESEARCH AND THE EXECUTIVE

Many "practical issues" limit the use of research in marketing. Some managerial problems are really ethical or value problems rather

¹⁹ The "classic" in this field is Walter A. Shewhart, *Economic Control of Quality of Manufactured Product* (New York: D. Van Nostrand Co., Inc., 1931). A recent text is Eugene L. Grant, *Statistical Quality Control* (rev. ed.; New York: McGraw-Hill Book Co., Inc., 1952).

than factual ones. For example, there might be a problem of deciding whether or not to market and promote a cold remedy in the face of good evidence that the remedy had no measurable impact on colds. For such value choices, research or any rational decision-making process is clearly irrelevant. But genuine value problems in marketing probably occur less frequently than generally believed.

"Science" and "research" are very often used as weapons rather than as tools for the analysis of problems. Often research serves to postpone the need for a decision to change policy, to "prove" points, to win arguments, and even to persuade people to buy the product. When research is used predominantly for persuasion, it is almost always used more or less dishonestly. The most subtle form of dishonesty and perhaps the most prevalent is the presentation of only part of the evidence and the suppression of the rest. Such dishonesty is often undetected but nonetheless futile simply because it is a weapon open to all parties. The prevalence of obvious disagreements between experts tends to reduce the layman's respect for research, and the suppression of unfavorable information is likely to become habitual.

Even when intentions are laudable, research is often aimless fact gathering that fails to provide help in making decisions except by accident. Some executives ignore research findings that are inconsistent with conclusions they have already reached; in the name of "evidence," some research men try to persuade executives to make decisions unsupported by the actual research. Perhaps the most revealing symptom of the practical difficulty of using research well is the frequent failure to define adequately the objectives of research. This failure probably stems from a lack of understanding as to what a full statement of objectives really involves and how much hard thinking it demands. Careful formulation of objectives need not and should not preclude the search for the unexpected "by-products" that are often so useful, nor does it preclude the collection of certain basic information—such as the rate of retail sales of one's product—in advance of current problems. But it is only too easy to lapse into the relatively mechanical assembly of information and quietly wander off into irrelevancy so that even useful information may not be properly understood by management even after sustained and aggressive attempts by research men to "sell" their product.

The failure to specify objectives adequately stems in part from human failings but also from the common tendency to regard marketing research as something essentially different from management or as a field for experts who supply the information needed for decisions to executives who then use intuitive judgment in arriving at the final decision. While there are potential advantages to be gained from specialization of functions, the intimate relationship between "research" and "intuition" together with the fact that the final decision rests with management puts on management the primary responsibility of using research well.

Marketing research is only one of the many fields of business in which experts in research aid executives. The executive's task is to use and evaluate the work of the expert without fully understanding the expert's discipline. The dependence on experts, of course, is not unique to business; the most unnerving example for many people is their own dependence on doctors. The executive's dependence on experts has also been recognized and sometimes deplored. The time, ability, and effort required to attain reasonable competence in any field of science ordinarily limits researchers themselves to relatively narrow specialization. Yet, the executive must make good use of the advice of many experts. To say that this task is extremely difficult is only to say that rational or even partly rational decisions are hard to make.

Perhaps the most important potential contribution of "research" in marketing comes from the relatively objective viewpoint that research encourages even among those who do not directly execute it. (The increasing orientation of management toward the "consumer point of view" is probably traceable to marketing research.) One cannot help being impressed by the frequency with which data contradict strongly held beliefs or suggest completely new alternatives for action. However difficult it may be to acquire the needed objectivity, the acquisition is probably not much more difficult for executives than for technically trained research people. In those areas of research where scientific experimentation is seldom used, of which marketing research is one, there are few occasions on which the research man is really "proved wrong," and it is being proved wrong rather than technical training *per se* that probably makes for genuine objectivity. Unfortunately, observational data usually can be interpreted in many plausible ways and frequently fail to convince either research men or executives that their previously held views were wrong.

One can be proved wrong also in the realm of logic, and the importance of sound logic in decision making should not be overlooked. Commonplace examples of importance are easy to find in practice: the assumption that the demand curve is infinitely inelastic for small price rises though not for large ones and the use of full costs rather than marginal costs in making decisions on output. While some errors of logic are extremely subtle and hard to detect, it would not seem unreasonable to expect that executives might learn to avoid the obvious ones. Mistakes of empirical inference also probably turn more frequently on elementary errors than on subtle ones.²⁰

Two prerequisites to more effective use of marketing research by executives can be embraced under "objectivity" and some understanding of basic tools of research—particularly logic, economics, and statistics. The fundamental key, however, is the executive's understanding that re-

²⁰ See Wallis and Roberts, *op. cit.*, chap. 3.

search, like management itself, must be evaluated ultimately by predictive tests. The single most useful step in fitting research into marketing management would be greater recognition of the fact that the essence of research is neither surveys nor samples but the evolution and testing of hypotheses about marketing by testing the predictions to which they lead.

There is, of course, much more to management than research. There are difficulties in applying research with which we have not attempted to deal, particularly the extent to which the scientific approach to management might dull intuition or lead to overintellectualization that could easily paralyze decisive action. Again, an executive who knows a certain venture to be a long shot may not try nearly so hard to make it succeed as if he believes, erroneously, that it is very likely to be successful. There is little which can be said about these things except to venture the opinion that they are probably overrated in importance. For each man who succeeds in apparently irrational ventures or whose intuition seems to surpass rational calculations, there are many for whom irrationality is the prime cause of failure.*

* In retrospect, many objectives only incompletely attained in this paper have either been attained, or brought within reach of attainment, by the development of Bayesian decision theory, a subject that was in its infancy in 1957. Those who are seriously interested in the problems raised in this paper will wish to study the subsequent literature of decision theory. One key reference is Robert Schlaifer, *Probability and Statistics for Business Decisions* (New York: McGraw-Hill Book Co., Inc., 1959). It is my current opinion that marketing research can be made far more useful and pertinent by learning how to apply decision-theoretic ideas.



FUNDAMENTALS

PART TWO



Research Design in Marketing Analysis*

.....

RONALD E. FRANK

INTRODUCTION

EVERY STUDY SHOULD HAVE A FRAMEWORK THAT WILL SERVE AS A BASIS for the collection and analysis of data in such a way that the study will be relevant to the problem and will use the most economical procedures. This framework will be called the research design. Research designs can be classified into three categories, according to the major purpose of the investigation. These are exploratory, descriptive, and causal. Exploratory designs are concerned with investigations whose purpose is to gain familiarity with a new phenomenon or to achieve new insights into it. For example, a study done by Social Research, Inc., for the *Chicago Tribune* on "Consumer Attitudes toward Beer and Beer Advertising," was designed to provide insights into the motivations underlying the purchase act for the purpose of developing hypotheses to serve as the basis for further research and as a basis for the development of campaign themes.

Descriptive designs usually are associated with two different types of problems. They are used to describe the characteristics of a particular situation or market, and they are used to determine the frequency with which something occurs or is associated with something else. In contrast to Social Research's exploratory study, the United States Brewers Foundation has sponsored a series of studies each of which provides a descrip-

* This paper is a substantial revision of an earlier technical note on research design (EA-M 455) used in the introductory marketing course at Harvard University, Graduate School of Business Administration.

The organization and the content of this paper follows closely that of Claire Sellitz, Marie Jahoda, Morton Deutsch, and Stuart Cook, *Research Methods in Social Relations* (rev. ed.; New York: Henry Holt and Co., Inc., 1960), pp. 50-143. The discussion of experimental design relies upon an article by John A. Howard and Harry V. Roberts, "Experimentation and Marketing Prediction," unpublished manuscript. Last, but not least, Hans Zeisel's *Say It with Figures* (rev. 4th ed.; New York: Harper & Bros., 1957) served as an excellent source of illustrations.

tion of the trends in beer consumption with respect to such characteristics as city size, age, sex, occupation, and geographic location, as well as such variables as social acceptance and consumption.

Causal designs, as the name implies, are concerned with the determination of cause and effect relationships. For example, a beer company might use different appeals in its local advertising in different markets in order to determine which appeal results in the highest rate of sales.

While the above classification of research designs is useful as a guide for gaining insight into the research process, it should not be taken too seriously in the case of any one study. Although the above examples have been classified as though their creators had only one of the three purposes in mind, a given design may, in practice, serve more than one purpose for a given investigation or may serve different purposes for different users. In spite of this fact, any specific research design is better suited to some tasks than to others. And a crucial tenet of research is that the design of an investigation should stem from the problem. In a sense, creating a research design is comparable to making a suit of clothes for a specific individual. The suit, during its lifetime, may be used by a number of individuals, but was made to fit one person.

In the following discussion, heavy emphasis will be given to those designs dealing with causal relationships¹ even though the majority of specific research projects in marketing are of an exploratory or a descriptive nature. There are two reasons for this emphasis:

1. From the standpoint of marketing management, the ultimate purpose in collecting research information is to reduce the risks of error in making decisions. Decision making implies the existence of a goal, as well as a set of alternatives as the basis for choice. The research process attempts to aid in determining which alternative act will maximize the decision maker's goal. The problem of determining the relationship between the decision maker's acts and his goal is simply a specific instance of the more general problem of isolating cause and effect relationships.

2. Where the purpose of a specific project is either exploratory or descriptive, the project is often the first step toward the evaluation of a possible change in the firm's marketing mix, though at times it may lead to direct action as opposed to further evaluation. For example, if the results of exploratory research with respect to a new product are favorable, the next step may be a sales test. Or, if it appears that a competitor is about to bring out a similar product, the exploratory research might be followed by a decision to introduce the product.

The decision orientation of the firm (and therefore its concern with consequences of alternative acts) should permeate its research activity with respect to both the ultimate purpose and design of research. Knowledge of the problems involved in the measurement of causal relations can

¹ The notion of a causal relationship will be discussed in more detail later.

be of use to marketing managers in their efforts to assess and design research in a way that will make it serve as a useful base for action.

The three types of research designs will be discussed in the following order:

1. Exploratory designs.
2. Descriptive designs.
3. Causal designs.

EXPLORATORY STUDIES

Many exploratory studies have the purpose of formulating a problem for more precise investigation or of developing hypotheses. An exploratory study may, however, have other functions: increasing the investigator's familiarity with the phenomenon he wishes to investigate in a subsequent, more highly structured study, or with the setting in which he plans to carry out such a study; clarifying concepts; establishing priorities for further research; gathering information about practical possibilities for carrying out research in real-life settings; providing a census of problems regarded as urgent by people working in a given field. . . .²

Exploratory or formulative studies are often seen as an *initial step* in a *continuous process*. By definition, when a researcher is at the *initial step* in the research process, he lacks a great deal of knowledge about the problem. If he were better informed and could develop specific hypotheses with respect to it, there would be no need to conduct an exploratory investigation. Because of this lack of familiarity with the subject matter, studies of this type are often designed in a flexible fashion that permits the investigator considerable freedom with respect to the methods used for gaining insight and developing hypotheses. This freedom is reflected by the fact that exploratory studies seldom use detailed questionnaires or involve probability sampling plans.

Emphasis on the *continuity* of the research process stresses the importance of exploratory research, which often becomes the basis for causal investigation and ultimately action. Looked at in this way exploratory research is an important determinant of what causal relationships are investigated, and therefore of the eventual actions taken by management.

Ingenuity, judgment, and good luck will inevitably play a part in determining a study's productiveness. It is nonetheless possible to suggest some methods that are likely to be especially fruitful in gaining the insight required to spot variables³ and set up meaningful hypotheses.⁴

² Selltiz, Jahoda, Deutsch, Cook, *op. cit.*, p. 51.

³ The word "variable" is defined by *Webster's New Collegiate Dictionary* as any magnitude that has different values under different conditions. Variables may be either quantitative or qualitative. Family income would be an example of the former, while geographic location (for example, farm versus rural) illustrates the other.

⁴ The word "hypothesis" is defined by *Webster's New Collegiate Dictionary* as a tentative theory or supposition provisionally adopted to explain certain facts and to guide in the investigation of others. An executive may observe that when his advertising budget is high the quantity of his product demanded is high and when the ad

These methods include: (1) a review of pertinent literature; (2) a survey of people who have had practical experience with the problem to be studied or with comparable problems; and (3) an analysis of "insight-stimulating" examples (as subsequently described). Most exploratory studies utilize one or more of these approaches. For example, suppose you were the research director of a firm considering entry into the brewing industry. In the early stages of this decision you might be interested in the present nature of the brewing industry. Just what kind of a process might you go through in attempting to gain insight into the nature of this industry? Obviously, one of the easiest things to do is to go to a library and look in the card index and at the *Business Periodicals Index* for books and articles on the brewing industry. In the process of reading a specific reference it is often worth noting the author's bibliography and in turn referring to his references. This "snowballing" technique provides a fast and efficient way of gaining information on a wide range of sources with relatively little effort. Reference to Standard and Poor's *Industry Surveys: Basic Analysis of the Liquor Industry* would provide data on such activities as prices, costs, inventory valuation, finances, and so on.

At some point it would probably be useful to take a look at the industry trade publications. Standard Rate and Data Service publishes advertising rate directories for media such as spot radio, spot television, consumer and farm magazines, and business publications. The main purpose of the publications is to provide data on the costs of advertising in business publications, however, they also provide a convenient listing of journals by industry. In addition, one could look at the *Encyclopedia of American Associations* for a list of trade organizations.

A glance at the *Monthly Catalog of United States Government Publications* might produce references on recent hearings or special studies. Frequently references such as these will provide intimate details of industry operation that normally do not appear in other published sources. Still another source of information consists of university research projects. *Dissertation Abstracts* provides a reference to a wide range of research activity. In addition, the Small Business Administration has published "A Survey of University Business and Economic Research Projects: 1957-1961."

At some point in the investigation, it would be useful to talk with people who have an intimate knowledge of the activities of the industry. The association directory, mentioned previously, provides a list of association officials. The *Standard Advertising Register* lists the location and the names of top officials for over 16,000 advertisers as well as their advertising agency and the account executive in charge.

budget is low, demand is low. This may lead him to hypothesize that there is a direct relationship between his advertising expenditures and the sales of his product. This hypothesis in turn might guide a more rigorous examination of this relationship.

In an exploratory investigation there is often a problem not only of finding information, but also of deciding what people and situations are apt to be the most fruitful bases for gaining insight. Though there are no simple rules for their selection, a few suggestions can be made:

1. Attention to changes, particularly abrupt changes, is often revealing with respect to the structure of an industry or a firm. For example, the adjustment of a market to the entrance of a new competitor can serve as a useful case history for developing hypotheses. How much of a market share does the new firm get? To what extent do the sales of the new firm represent an increase in industry demand as opposed to the switching of customers from one competitor to another? To the extent that sales represent switching, from which firms do the new customers come? Questions such as these, focused on a situation in which an abrupt change has occurred, can often give insight into the underlying competitive conditions in an industry.

2. Reviewing extremes of behavior can also be a useful device. For example, one might study the firms which have shown the most and the least growth during the last several years. One might interview two groups of executives, with comparable experience in the industry except for the fact that one group is considered "successful" by the trade and the other "unsuccessful."

3. Studying different positions in the structure of an industry also can be used to provide insight. One might talk, not only to beer producers, but also to distributors and retailers as well as to trade associations and regulatory bodies. Each of these groups is apt to have a somewhat different viewpoint.

4. Observing the order in which events occur over time is another device which might provide some clues as to the factors that are associated with different levels of performance in the industry.

The methods of research and thought discussed above often do not provide evidence on or from "typical" sources. The idea is to try to find those situations or persons who are apt to be the most fruitful from the standpoint of gaining insight.

DESCRIPTIVE STUDIES

A great deal of research has concerned the characteristics of some aspect of a marketing function or problem. There are studies which describe the demographic characteristics of the consumer (for example, percentage distributions of age, sex, education, location); the location of a firm's salesmen, wholesalers, or retailers; the changes over time in the cost and profit structure of a firm or an industry; the characteristics of successful salesmen; or the association between amount of a product consumed by a family and its geographic location. In each of the above cases the purpose of the investigation was to describe the distribution of some single phenomenon (for example, the percentage distribution of consumers by age) or to describe the relation between two or more phenomena (for example, the relation between the amount consumed by a customer and where he lives).

There are a number of decision-making situations where descriptive information can serve as a partial basis for action. For example, suppose you are entering a new market area. Suppose further that past research

and your own experience have led you to believe strongly that sales of your product are closely associated with family income. In this new market, it may be sufficient for the research department to tell you the location and number of high income families. As a second illustration, you are about to introduce a product which is superior to previous products of the same type but costs half as much. At the outset you may be mainly interested in characteristics of present consumers of the product so that you will know where to distribute and advertise the innovation, as opposed to being immediately interested in why consumption in Kansas is higher than in California.

Both descriptive and causal designs presuppose prior knowledge of the subject to be investigated, as contrasted with the questions that form the basis for exploratory studies. There is a need to define clearly what is to be measured and to find adequate methods for measuring it. In addition, care must be taken to specify what is to be included in the definition of "a given market" or "a given population." What is needed is a clear formulation of what and who is to be measured and techniques for valid and reliable measurements, and not so much the flexibility characteristic of an exploratory study.

The need for data collection to be based upon a clear-cut definition of the problem is illustrated in the following situation: Suppose your firm is considering whether or not to build a plant at location A. What factors would help you make this decision?

One item might be an evaluation of the firm's profitability, conditional upon the expected sales volume. Your firm has kept records of past sales by geographic areas for use in evaluating the various elements of the marketing mix. Could you just add up the past sales from the area to be supplied by the new plant, and use this for the basis of estimating its profitability?

Are the assumptions implicit in using these routinely collected data reasonable for the problem at hand? They may be. But they often warrant checking. For example, will the building of a new plant in a specific region change the competitive situation? How will your competitors react? Will one build a plant in the same area? If such retaliation seems reasonable, what effects would it have on your sales forecast? Data collected for general purposes are apt not to reflect the idiosyncratic characteristics of individual decisions.

The need for care in defining the problem and for the detailed specification of the data to be collected are the main characteristics of descriptive research. Causal research, in contrast, not only requires the care and precision demanded in a descriptive study, but also requires procedures that will permit inferences about causality.

CAUSAL STUDIES

The discussion will first focus upon the nature of causal relationships, and then upon the bases for inferring their direction and magnitude. This

will be followed by a brief discussion of alternative types of research designs aimed at measuring causation. Next will be a discussion of the principles of experimental design as they apply to the problem of inferring the direction and magnitude of a causal relationship. The next section will contrast the techniques utilized in experimental designs with those used in nonexperimental research. This contrast will serve as a basis for gaining a clearer understanding of experimentation and observation, and for coordinating the material presented in this section with what is to follow in the discussion on statistical analysis covered in the paper to follow. The paper will conclude with a discussion of the limitations of experimental designs and the presentation of an outline for the evaluation of experiments.

A causal hypothesis asserts that if a particular characteristic (for example, number of retail outlets) is varied while all other characteristics are held constant, the criterion (for example, profits) will vary in some particular manner. This hypothesis can be looked upon as making two assertions: (1) that the number of retail outlets and the magnitude of profits in an area are related to each other and (2) that the relationship runs in one direction, from the retail outlets to profits. In other words, causal hypotheses are asymmetrical. They are defined as relations that go in one direction only. From the standpoint of inferring a causal relationship it is not enough to simply observe an association (for example, the higher the number of retail outlets our firm has in an area the higher our profits). In addition, one must develop some bases or rationale for inferring that the relationship was asymmetrical,⁵ such as data on overtime relationships or associations at a point in time. Cause can never be observed. It can only be inferred.

The discussion that follows will be aimed at answering the following questions:

1. What bases can be used for inferring a causal relationship?
2. What are the types of research designs that are used for measuring causal relations?
3. What are the characteristics of each of the designs, and how do these compare as a basis for measuring the direction and magnitude of the relationship?
4. What criteria can be used for evaluating causal designs and subsequent results as a partial basis for decisions?

Questions of substance in marketing frequently take the form of:

1. What is the effect of changing the level of X (the brand's price) on Y (the firm's market share)?
2. Given the firm's present X_1 (product quality) and X_2 (advertising budget), how should an additional expenditure of \$100,000 (aimed at a

⁵ For a more detailed discussion of the notion of causation see Herbert A. Simon, "Causal Ordering and Identifiability," in William C. Hood and Tjalling C. Koopmans (eds.), *Studies in Econometric Method* (New York: John Wiley & Sons, Inc., 1953), pp. 49-74.

change in the firm's product quality or ad budget) be allocated in order to achieve the greatest increase in $\bar{Y}$ (profits)?

In marketing, as well as in the behavioral sciences, a multiplicity of conditions, taken together, make an event probable. One seldom, if ever, finds a situation where a single cause always leads to another single effect. Management, in the hope of decreasing errors with respect to its decisions (and thus reducing opportunity loss⁶) faces the problem of measuring the probable effects of each course of action. What types of evidence can be used as a basis for the measurement of causation?

BASES FOR DETERMINING THE NATURE AND EXTENT OF A CAUSAL RELATIONSHIP

There are three types of evidence that serve as the basis for evaluating causal relationships. They are concomitant variation, time order of occurrence of variables, and the elimination of other possible causal variables.

Concomitant Variation

Concomitant variation *between two variables* (say, X and Y) is defined as the association between the values of one variable and those of another. If X is one of the causes of Y , we would usually expect to find a greater proportion of our subjects with Y , given X , than without X .

Concomitant variation *between the changes in two variables* (say, ΔX and ΔY) is defined as the association between the magnitude of the change in one variable and that in another. If ΔX is a cause of ΔY then one would expect the magnitude of ΔY to be different in situations where ΔX was large as opposed to situations where ΔX was relatively small.

For example, suppose a divisional manager of a food chain noticed that per customer sales of apples had increased during the past week, and that a number of stores were using window displays of apples. How would the divisional manager find out the direction and magnitude of the relationship between sales increases and window displays?

One way would be to sort stores by the magnitude of their per customer apple sales to see if stores with high sales volumes tended to use window displays to a greater degree than those with lesser volumes.

A second way would be to compare stores with large sales increases with those having a small increase or a decrease. If none of the stores with a large increase had window displays, he would conclude that the latter were not the cause of the sales increase. Suppose some stores showing large increases in apple sales had used window displays, while others

⁶The opportunity loss of a decision is defined as "the difference between the cost or profit *actually* realized under that decision and the cost or profit which *would have been* realized if the decision had been the best one possible for the event which actually occurred," in Robert Schlaifer, *Probability and Statistics for Business Decisions* (New York: McGraw-Hill Book Co., Inc., 1959), p. 117.

had not. What then? He might look at stores whose per customer sales had decreased to see what fraction of them had displays. If the displays were one of the causes of the increase in per customer sales of apples, he would expect stores with a large increase in sales to have a higher proportion of window display users than stores whose sales performance had been in the opposite direction.

In contrast to these, another way of getting evidence on the same subject would be to see if either window display users tended to be stores with higher per customer apple sales than nonusers or if window display users tended to have higher sales increases than nonusers.

Two approaches to the determination of concomitant variation have been illustrated. The first starts with the variable you want to effect, often called the dependent variable (increased sales), and then goes back to the variable which you think may help to determine the behavior of the dependent variable (called the independent variable). The second starts with the independent variable (window displays) and attempts to determine its relationship to the dependent variable. The use of the second strategy suggests that the researcher knows what causal relation he wants to test, while the first is more apt to be associated with a "fishing expedition" aimed at finding out what the causes of a sales increase were without specifying in advance just which hypotheses might be tested.

In the preceding paragraphs concomitant variation has been defined in terms of the levels of two variables and in terms of changes in the levels of two variables. Whether or not you want to use levels or changes in levels depends, in large part, on the nature of the problem and the data with which you are concerned. In the case of the window display and apple sales example, it may be that stores with high sales volumes use more displays when they are available because they have a larger market for apples, than those with lower sales volumes. A comparison of the level of apple sales among these two groups of stores would lead to an overstatement of the effect of window displays on apple sales. One way of trying to get around this problem would be to look at changes in the sales volume of a store from one period to the next for those stores where window displays were introduced in the latter period versus those where no window displays were used. In other words, it might be more reasonable to assume that a comparison of the magnitude of the change from one period to the next would be less effected by the differential behavior of small and large stores than would a comparison of the levels.⁷

Based on evidence of concomitant variation, one might be tempted to conclude that window displays lead to increased sales. Unfortunately, the

⁷ Still another possibility would be to measure *percentage* changes from one period to the next rather than the *absolute* changes. This model might be superior to the absolute change model if one believed (or found) that stores with larger sales volumes tend to have period to period changes of a greater magnitude than those with smaller volumes.

extent (and even the direction) of an association between X and Y cannot be taken as a measure of the extent and direction of causation in and of itself. For example, in a study of candy consumption⁸ it was found that a higher percentage of single as opposed to married people eat candy regularly. From this result could one safely infer that marriage causes a decrease in candy consumption? Or is there another possible explanation? How about the effects of age? Married people are older than single people, and perhaps older people eat less candy. An analysis of the effect of age revealed that the relationship between marriage and candy consumption was almost completely explained by the fact that married people tended to be older (that is, if one looks at people of the same age their candy consumption would be the same regardless of whether or not they were married). The relationship between marriage and candy consumption was due to a third variable, one that was correlated with each of the other two variables.⁹

On the other hand, the virtual absence of association cannot be taken as evidence, in and of itself, of the virtual absence of a causal relationship. For example, Zeisel reports a study by Lazarsfeld¹⁰ in which the percentage of individuals listening to classical music was found to be virtually the same regardless of the age of the respondents. This result was contrary to expectations. However, when the respondents were classified by educational level, it was found that listening to classical music increased with age among respondents with a high educational level and decreased with age among those with a low educational level. When the high and low educated groups were combined, the effects of these two tendencies compensated for each other, and resulted in a nearly complete lack of association between age and listening to classical music.

In both the candy and music illustrations, association per se did not constitute a causal relationship. In each case, the analyst's knowledge of the problem under study, combined with his knowledge of technique, led to a refinement of the original results. The original results seemed "unreasonable" in the context of the problem. The analyst could find no rationale consistent with his own experiences that would explain the original relationship. This led him to look for other possibilities. The analyst's knowledge of the subject under study played an important role in shaping his interpretation of the results.

As the district manager faced with the problem of evaluating the effect of window displays on apple sales, would you be satisfied with the evidence of concomitant variation described previously? Why not? What

⁸ This example is reported by Hans Zeisel, *Say It with Figures* (rev. 4th ed.; New York: Harper & Bros., 1957), pp. 197-200. This text is highly recommended to persons interested in a concise, easy-to-read introduction to the problems of analyzing and evaluating information.

⁹ This point will be discussed further in the following paper by William Massy.

¹⁰ *Ibid.*, pp. 181-82, adapted from Paul F. Lazarsfeld, *Radio and The Printed Page* (New York: Duell, Sloan and Pearce, 1940), p. 80.

other factors might account for the observed relationship? To what extent do you think your knowledge (or absence of knowledge, as the case may be) with respect to the problems of in-store promotion affects your answer to this question?

Time Order of Occurrence of Changes in Behavior

Uncertainty with respect to the direction of a relationship can often be reduced by observing the time order in which changes in the variables occurred. If the increase in per customer apple sales occurred before the displays were put up, it could hardly be argued that the window displays caused higher sales. Or, if there had not been a change in display during the period under study (that is, if stores with no display during the first period had none during the second, while stores with displays had kept them), then the sales increases could not be directly attributable to the use of displays.

Not all temporal relationships have obvious interpretations. For example, the relationship between changes in annual advertising expenditures and changes in sales for a given product is frequently used as evidence of the effect of advertising on sales. Many firms use a rule of thumb such as 10% of sales for advertising. The question arises as to which way the relationship runs. Does increased advertising lead to higher sales, or do higher sales lead to an increased advertising budget? The analyst's knowledge of the decision rules used for setting the advertising budget of the firm, would shed light on the interpretation of this relationship in a given situation.

One of the most graphic illustrations of the ambiguity of over time relationships, taken as evidence in and of themselves, concerns the relationship between stock prices and dividends. Suppose a man from Mars completely ignorant of our society landed on earth and observed that stock prices rose just before the payment of dividends. He might infer that increases in stock prices lead to the payment of dividends. However, a knowledgeable citizen from planet earth would place a considerably different interpretation as to the direction of causation. His rationale for the observed relationship might include the fact that stock prices rise in anticipation of dividend payments and that the principal effect is one of dividends on prices.

Our man from Mars would be faced with a similar problem if he observed the over time relationship of family consumption to income (particularly for upward mobile college graduates). Increases in consumption would appear to cause increased income (one can contemplate some situations where this would be true). In a fashion similar to stock prices, anticipation of future income increases is apt to lead to increases in family consumption.

Once again the importance of the analyst's knowledge of the problem area plays an important role in the interpretation of the observed relationship.

Elimination of Other Possible Causal Factors

In the context of the problem of evaluating the effect of window displays on apple sales, several alternative causes of the sales increase might suggest themselves to the analyst. The same stores that used window displays of apples may also have had a special price or special displays within the stores. The increase in sales per customer might be due primarily to these promotional activities as opposed to the use of a window display. One might compare stores simultaneously as to whether or not they used a window display and as to whether or not they used a special in-store display to determine the extent to which each of these factors accounted for the sales increase. It is often possible to build into the design of a study ways of controlling a few of the variables whose effects might lead to an erroneous interpretation.

Another device frequently used for making the effects of other variables more predictable is that of matching. Just what matching is, and its use to achieve this goal, will be discussed in a subsequent section of this paper.

Each of the three types of evidence that have been discussed helps to serve as a basis for evaluating the existence of a causal relationship. Such evidence merely provides a reasonable basis for determining the direction and magnitude of the relationship between X and Y . There is always the possibility that some factor has been left out of the analysis, which if taken into account, would change either the magnitude or both the direction and magnitude of the relationship. One may conclude that it is reasonable to believe that the magnitude is roughly of a certain degree, but one can never *conclusively* demonstrate this.

In attempting to evaluate a particular observed relationship, one frequently looks for a series of pieces of information that form a pattern as opposed to relying on any single piece of evidence as the basis for measuring the direction and magnitude of a relationship.¹¹ For example, in attempting to evaluate the effect of window displays on apple sales, we have illustrated how each of the three types of evidence can be employed. Researchers frequently will attempt to build ways of using all three types of evidence into a study so as to have as many checks as possible as to the validity of their findings. This pattern of evidence together with an intimate knowledge of the problem under study, serves as the basis for interpreting the results of a research.

TYPES OF RESEARCH DESIGNS AIMED AT INFERRING CAUSATION

Research designs vary greatly in terms of their adequacy as a basis for inferring the existence of a causal relationship. Designs can be divided

¹¹ Techniques for statistically taking into account the effect of other variables, as well as for simultaneously analyzing the relevance of several different types of evidence with respect to predicting behavior (multivariate analysis) are discussed in the following paper: William Massy, *Statistical Analysis of Relations between Variables*.

into two major categories. They are experimentation and observation.¹² Experiments are defined as studies in which implementation involves intervention by the observer beyond that required for measurement, whereas observational studies involve only that degree of intervention required for measurement.

Our window display example can serve as an illustration of an observational study. Store managers were free to choose whether or not they would use a window display (that is, there was no intervention by the observer). An attempt can be made to contrast the sales increase of stores with window displays against those without displays.

We could change this to an experiment by requiring all managers to use displays during some specified period of time. Our experiment might then be used to measure the effect of the displays in terms of the increase in per customer apple sales from the week prior to the intervention to the time of the intervention. This form of an experiment is often called a "tryout."

One question left unanswered by the tryout is whether or not sales would have increased from week to week if no deliberate change had been made. Our tryout could be modified in order to take into account this possibility. We could divide our stores into two groups. One group would use displays. The other group would continue as they had been with no change in conditions brought about by the experiment. This group is usually known as a control group.

The usefulness of such a control group is easy to illustrate. Given the division of stores into control and test groups, we then introduce window displays into the test group. If sales double in the test group, management's response to this doubling should be affected by the results in the *control group*. If the control group's sales actually decreased, it would be much more reasonable to attribute the increase in sales in the test group to the use of window displays than if the control group's sales also doubled. A doubling of the control group's sales, on the other hand, would seem to imply that some other factor or combination of factors caused the increase.

By specifying what stores are to be included in the experimental and control groups, another source of possible bias inherent in our observational study can be reduced. In the observational study, store managers decided for themselves whether or not to use window displays. The fact that sales increased in stores with window displays could be attributable to the ability of store managers to discern under what conditions window displays will be effective (for example, managers might have observed that displays tend to be more effective in high income neighborhoods or

¹² The use of the term "observation" in this context does not mean to imply that observational studies are concerned with only the problem of measuring causal relations. Both the exploratory and descriptive studies can be thought of as types of observational studies.

when prices of competing fruits are high relative to apples). If the researcher's problem is to evaluate the use of window displays, he needs a method for dividing stores into groups that will avoid the possible "self-selection" bias of the managers. The principal techniques used for such grouping are randomization and matching. The major purpose of randomizing and matching is to help insure that the different groups in the experiment are comparable before intervention, except for variations that can be predicted by chance.

By its use of intervention, a control group, and techniques of randomization and matching, the experimental method provides the most effective method of measuring causal relationships. Despite this fact, there is always the possibility of a fallacious inference. Confidence in the results of research should therefore depend not only on the sophistication of the research design, but also on the manager's previous experience and his knowledge of past research.

Despite the methodological advantages inherent in intervention and the use of control groups, observational studies are still the most frequently used basis for inferring causal relations. Two of the principal reasons for this fact are: (1) that experimentation is economically infeasible with respect to many problems and (2) that it is often difficult to intervene in the desired fashion. An example of the first limitation would be the study of the effect of industry structure on price flexibility (one could hardly modify structure just to make the study). The second limitation would occur in situations concerned with media evaluation (for example, magazines) where one usually cannot blank out a single city in and of itself.¹³

There are three types of observational studies: (1) time series analysis, (2) cross-sectional studies (studies at a point in time), and (3) studies which combine both time series and point in time analysis.¹⁴ While observational studies do not involve intervention and randomization, a number of substitute techniques have been developed to provide the same sort of safeguards against unwarranted conclusions, which will be described in detail shortly.

In the next section a detailed discussion of the use of experiments as a basis for estimating causal relationships will be presented. This will be followed by a section on observational studies. The paper will conclude with a brief discussion of some of the factors that limit the more widespread use of experimentation, thereby leading to the frequent use of observational designs. Major emphasis is given to the principles of experimentation, not because experimentation is the most common approach in marketing research (which it is not), but because its principles provide

¹³ Though for some magazines one can blank out a region.

¹⁴ To be described in detail in a subsequent section of the paper.

an ideal model for the design of both experimental and observational studies.

THE MEASUREMENT OF CAUSAL RELATIONSHIPS: EXPERIMENTATION

The design of an experiment makes possible the collection of the three major types of evidence relevant to evaluating causal relationships:

1. Evidence with respect to concomitant variation.
2. Evidence with respect to the time order of variables.
3. Evidence ruling out other variables as possible causes.

Evidence of the first type is provided by either observing how much *sales* or the *increase in sales* is greater among subjects who have been exposed to window displays than among those who have not been exposed.

The second type of evidence is provided by setting up control and experimental groups in such a fashion that it is reasonable to assume they did not differ in terms of the dependent variable and/or by measuring the differences between groups with respect to the dependent variable before exposure to the variable under study. If the dependent variable were sales, one might measure the sales volume of the experimental and control groups before the study. If the volume of sales for the experimental groups were twice that of the control before intervention, knowing this fact would materially affect one's interpretation of a ratio of two to one between the experimental and control groups after intervention.

The nature of evidence concerning the ruling out of other factors depends upon the factor involved. Some of the major classes of other variables are:

1. Enduring characteristics of the subjects.
2. The effect of the measuring process itself.
3. Unpredictable events that occur during the course of the study.
4. Gradual changes that occur over time.

Illustrations of types one and two are discussed in subsequent sections of the paper. With respect to unpredictable events, suppose one were testing the effect of an advertising campaign on sugar sales, and suddenly there were a war scare. The chances of still having useful results would be somewhat better with an experiment because, hopefully, the effect of this event would be the same on both the experimental and control groups, thereby still allowing a basis for comparison. In a similar fashion in an experiment, the effect of a gradual change such as a seasonal fluctuation in sales stands a better chance of being adjusted for with the use of a control group.

Three aspects of experimental design are particularly important from the standpoint of protecting against unwarranted inferences with respect

to cause and effect relationships. These are the selection of the experimental and control groups, the way in which control groups are used, and the number of possible causal variables that can be systematically included in a study. Each of these will be discussed in the sections that follow.

Types of Experimental Design

Experiments may be classified according to: (1) whether "before" as well as "after" measurements are taken, and (2) the pattern of control groups used. It will be shown that under some conditions inappropriate choice of control groups or use of after-only measurement will materially increase the chances of misinterpreting the results of an investigation. Therefore, both of these aspects must be carefully considered in the design and evaluation of a study. The discussion that follows will present examples of several experimental designs, aimed at illustrating variations in the use of control groups, as well as the use of before and after measurements.¹⁵

After-Only Design. The format for an after-only study with one control group is as follows:

	<i>Experimental Group</i>	<i>Control Group</i>
Prior selection	Yes	Yes
Before measurement	No	No
Experimental variable	Yes	No
Uncontrolled events	Yes	Yes
After measurement	Yes (Y_1)	Yes (Y_2)

Prior selection of the groups is based on matching and randomization techniques. Matching constitutes the chief protection for assuring that the groups are comparable. No before measurement is taken. The experimental variable is introduced and its effect (E) is measured by $Y_1 - Y_2$. This design avoids the problem of interaction¹⁶ and is apt to cost less to administer than a before-after design.

One of the most frequent applications of experimentation has been with respect to the evaluation of direct mail campaigns. For example, suppose you were faced with the problem of choosing one of three direct mail pieces for use in a forthcoming mail order campaign. We could take a sample of households from the mailing list to be used as the basis for the mailing. Using appropriate techniques of matching and/or randomization, we could separate the sample into three groups.¹⁷ Each

¹⁵ For a more detailed treatment of the subject see D. R. Cox, *Planning of Experiments* (New York: John Wiley & Sons, Inc., 1958); William G. Cochran, and Gertrude M. Cox, *Experimental Designs* (2d ed.; New York: John Wiley & Sons, Inc., 1957); and Walter T. Federer, *Experimental Design* (New York: The Macmillan Co., 1955).

¹⁶ To be discussed in the following section.

¹⁷ There are conditions under which it may not be feasible to randomize. These are discussed in a later section of this paper which is concerned with the limitations of experimentation.

group would receive a different direct mail piece. The orders resulting from the test mailing would then serve as the basis for comparing the relative effectiveness of each of the three pieces.

For problems where management wants to choose one of a set of alternatives, such as in ad selection or direct mail evaluation, the after-only design can be quite useful. However, while the techniques of matching and randomization may have been used to assure the comparability of the groups before intervention, some differences may still exist. With the after-only design the researcher has no measure of their magnitude.

Before-After with One Control Group. This design may be visualized as follows:

	<i>Experimental Group</i>	<i>Control Group</i>
Prior selection.....	Yes	Yes
Before measurement.....	Yes (Y_1)	Yes (Y_1')
Experimental variable.....	Yes	No
Uncontrolled events.....	Yes	Yes
After measurement.....	Yes (Y_2)	Yes (Y_2')

The plan is aimed at measuring the effect of one variable. The experimental and control groups are selected in a fashion aimed at insuring comparability. Next, before measurements are made of both groups. The experimental variable is then introduced. Both groups are exposed to uncontrolled conditions, and after measurements are subsequently made. The difference between the after and before measurements in the control groups ($Y_2' - Y_1'$) reflects only the effect of uncontrolled variables, while the difference between after and before measurements in the experimental group reflects the effect of the experimental variable and the uncontrolled factors ($Y_2 - Y_1$). Therefore, the effect of the experimental variable alone is shown by the difference between the effects in the experimental and control groups [$(Y_2 - Y_1) - (Y_2' - Y_1')$].

Our window display test could be designed in this fashion. The before measurement would correspond to per customer apple sales during the week prior to the test. The experimental variable would be the window displays. Both the experimental and control groups would be exposed to uncontrolled events such as variations in the weather or changes in competitive conditions. The after measurement would be per customer apple sales during the period the window displays were in use. While the logic of this model is sound,¹⁸ its usefulness can be seriously impaired in situations where whatever is being measured can be affected through

¹⁸ The appropriateness of this measure of the effect of window displays depends in part on the economic interests of the decision maker. This measure of increased revenue due to displays might serve as a useful basis for action from the standpoint of a trade association interested in promoting apple sales. However, this same measure may be inadequate from the standpoint of the chain store operator who might also be interested in the revenue lost due to the substitution of apple purchases for those of other commodities.

the process of measurement itself. In the window display test, if the taking of before measurements alerted the store managers in the experimental group causing them "to take advantage" of the situation by increasing their in-store promotional activity during the tests, then the effect $Y_2 - Y_1$ would be overstated. If at the same time members of the control group tended to decrease their promotional efforts on apples due to the fact that they would be unable to use window displays during the test period, then the effect $Y_2' - Y_1'$ would be understated. The net effect of these two conditions would be to overstate the effect of the experimental variable $[(Y_2 - Y_1) - (Y_2' - Y_1')]$.

Another example of the effect of the before measurement was reported in a study of a United Nations educational campaign.¹⁹ Two equivalent samples of 1,000 people were selected. A before measurement of information and attitude toward the United Nations was taken from only one of these samples. The other sample was interviewed after the educational campaign. The members of this sample were neither better informed nor did they have different opinions than the members of the first sample. The campaign apparently had no effect. The members of the first sample were then reinterviewed. This group had undergone definite changes in attitude and information about the United Nations. In this situation, the taking of a before measurement made people more aware of the subject and more apt to be influenced by the advertising.

The important point for our purpose is that in situations dealing with human beings the results of a single before-after design can be distorted by interaction between the before measurement and the experimental variable. The effect of the experimental variable may be different depending on whether or not a before measurement was taken.²⁰ Two ways of getting around this problem would be to use an after-only study or to use a before-after design with two control groups.

Before-After with Two Control Groups. One method of overcoming the problem of interaction between the before measurement and the experimental variable is to modify the pattern of control groups, as is illustrated in the following design:

	Experimental Group	Control Group I	Control Group II
Prior selection.....	Yes	Yes	Yes
Before measurement.....	Yes (Y_1)	Yes (Y_1')	No ($Y_1'' = \frac{Y_1 + Y_1'}{2}$)
Experimental variable.....	Yes	No	Yes
Uncontrolled events.....	Yes	Yes	Yes
After measurement.....	Yes (Y_2)	Yes (Y_2')	Yes (Y_2'')

¹⁹ S. A. Star and H. M. Hughes, "Report on an Educational Campaign, The Cincinnati Plan for the United Nations," *American Journal of Sociology*, Vol. LV (1949-50), p. 389.

²⁰ We shall return to this point in a later section, when we will discuss panel studies.

The measurements made of the experimental group are the same as those made when only one control is involved. Again, $Y_2 - Y_1$ represents the difference between the after and before measurements for the experimental group. The pattern of the control groups, however, has been modified so as to separate the effect of measurement from the effect of the experimental variable.

The difference between Y_2 and Y_1 can be thought of as being composed of the sum of several components; namely the experimental variable (E), uncontrolled events (U), and the interaction of the before measurement with experimental variable (I). In other words, $Y_2 - Y_1 = E + U + I$. The first control group measures the effect of the uncontrolled variables only (that is, $Y_2' - Y_1' = U$). When compared with the experimental group, it shows the combined effect of the experimental variable and interaction:

$$(Y_2 - Y_1) - (Y_2' - Y_1') = (E + I + U) - U = E + I.$$

The objective of the second control group is to separate the effect of interaction from that of the experimental variable. While no actual before measurements are taken of this group, it would seem reasonable to assume (if an appropriate technique of selection has been used) that the results of such a measurement would be equivalent to those for the other two groups. Y_1'' therefore, consists of an average of the before measurements for the experimental and the control groups. Because no before measurement was taken there is no possibility of interaction occurring and therefore the difference between Y_2'' and Y_1'' is due to the experimental variable and uncontrolled conditions $Y_2'' - Y_1'' = E + U$. By appropriately matching each of the following experimental and control groups it is possible to make separate estimates of the effect of the experimental variable and of interaction:²¹

$$\begin{aligned} Y_2 - Y_1 &= E + I + U \\ Y_2' - Y_1' &= U \\ Y_2'' - Y_1'' &= E + U \end{aligned}$$

Along with the ability to separate out the effects of the measuring process from the experimental variable comes the problem of increased complexity and expense of the research design. Fortunately in some marketing studies there is no need for the subjects to know that they are being measured, and therefore, no danger of the measuring process impairing the results of the study. If store sales or sales by product are col-

²¹ One source of interaction that this design (before-after with two control groups) does not adjust for, is that between the before and after measurement (that is, once a before measurement has been taken it can, in and of itself, affect the magnitude of the after measurement). How might the design be modified so as to allow for the estimation of the effect of interaction between the before and after measurements (I')? With I' added we can restate the result of the change in the experimental group as the following sum: $E + I + I' + U$.

lected as a matter of routine, no special request would have to be made for the purpose of using them as a measure of performance in an experiment.

The preceding section has demonstrated the effect of changing either the pattern of control groups or the use of before and after measurements on the usefulness of a research design in the context of different problems. The section to follow concentrates on the problems involved in separating the units to be observed into experimental and control groups.

Selecting the Experimental and Control Groups

It is impossible to design an experiment that will tell us with complete certainty what the effect of the window displays has been on per customer sales. But it is possible to come up with a result which will allow us to estimate the probability that the actual effect has been within a given range. (Given the result we can make statements such as: the probability of the window displays increasing per customer sales by \$0.50 to \$1.00 is 0.4.)

The executive's ability to make such probabilistic statements depends on two things:

1. His past experience.
2. The extent to which statistical decision theory can be brought to bear on his problem.

In many situations, statistical decision theory can serve as a mechanism to combine the executive's prior beliefs with the results of a study. However, in order to be able to get much out of this tool for interpretation of an experiment, certain conditions must be met in the course of the design of the experiment. Before introducing the experimental variable, something needs to be done in order to insure that the experimental and control groups are sufficiently comparable so that measurements of the effect of the experimental variable can be made efficiently. That something is matching.

In the example on the relationship of window displays to apple sales such variables as the time of day and the day of week would have an effect on per customer apple sales. In designing a study of the effect of window displays one might match stores according to these characteristics. Suppose there will be 50 stores included in the experiment. One procedure for matching them for time-of-day and day-of-week would be as follows: (1) Divide the stores into five groups of 10 stores each. Each group of 10 stores will participate in the experiment on a given weekday; and then (2) Divide each group of 10 stores into five pairs of two stores each. Each pair of stores will participate in the experiment at a given time of day. The above procedure results in a design with five matched pairs of stores for each day of the week, five for each of five times of the day, and one for each day-of-week and time-of-day combined.

Once the matching process has been completed, randomization may be used to assign the members of each pair of stores to their respective groups. How do we separate them into experimental groups and control groups? We could flip a coin, for a store in each pair, with heads corresponding to the experimental group and tails the control group.²² While this technique of randomization does not guarantee that the experimental and control groups are identical, it is a method which assures us that the differences that exist before the introduction of the experimental variable are due to chance alone. The rules of statistical inference, together with an executive's judgment and prior experience can then be used more efficiently in interpreting the results of the experiment. Randomization must be used with care. For small sample sizes, randomization is unwise.

Consider the introduction of a new product in a number of test cities. Suppose that over-riding economic reasons dictate that at most three test cities can be chosen. Should they be picked at random? Here is an actual example: Three matched pairs of cities were chosen randomly, and a city in each of the three pairs was selected randomly as a test city; the other was a control. Use of the product was closely related to latitude. It was found that in each of the three pairs, the test city was south of the control city by about 150 miles. The original random choice was then modified to compensate for this obvious source of error.²³

Matching can be extended to embrace more than one variable at a time. For example, suppose, in addition to day-of-week, treatments were also to be matched with respect to time-of-day. Suppose an after-only experiment is to be performed with respect to the effect of window displays on per customer apple sales. If either day-of-week, or time-of-day are related to per customer apple sales, it is apt to be advantageous to see that the different groups of experimental units for each treatment, have similar distributions with respect to day-of-week and time-of-day (that is, are "balanced" with respect to day-of-week and time-of-day). One example of an experimental design aimed at balancing more than one variable (day-of-week and time-of-day), in order to get a more precise measure of the effect of the third (window displays) is the latin square. Suppose we wish to test four displays (treatments), while balancing day-of-week and time-of-day. Let us label by the four treatments T_1 , T_2 , T_3 , T_4 . The following design, a 4×4 latin square,²⁴ achieves our objective:

²² Experimental and control groups need not be (and frequently are not) the same size. For example, one could have groups of five stores matched with respect to some characteristic, assign each store a random number, arrange the random numbers in each group of five stores by magnitude, and assign the three stores in each group with the highest numbers to the experimental group. This would result in a design with 60 percent of the stores in the experimental group.

²³ Howard and Roberts, *op. cit.*, p. 15.

²⁴ In a latin square design, the number of treatments must equal the number of rows and columns. The most commonly used range is from 4×4 to 8×8 . Small squares, the 2×2 and 3×3 , are seldom used because of their relative lack of precision.

	<i>Day 1</i>	<i>Day 2</i>	<i>Day 3</i>	<i>Day 4</i>
Early morning.....	T_1	T_2	T_3	T_4
Late morning.....	T_4	T_1	T_2	T_3
Early afternoon.....	T_3	T_4	T_1	T_2
Late afternoon.....	T_2	T_3	T_4	T_1

Treatment 1 is administered to one store (or group of stores) on day 1 in early morning. Treatment 1 is also administered in the late-morning of day 2. Treatment 1 occurs once during each day of the week and once during each time of day. The same is true for each of the other treatments. If one is interested in whether or not treatment 1 outperformed treatment 2, he can compare the average performance for treatment 1 against the average for treatment 2. The performance with treatment 1 can be thought of as a weighted average, where the weights given to each day of-week or time-of-day are equal. Treatment 2's performance is based on experimental units which are averaged together using the same set of weights with respect to day-of-week and time-of-day as were used with treatment 1. The same is true for treatments 3 and 4. In a design where this was not done, variations in the weight given different days of the week or different times of the day would be uncontrolled. By controlling this variation, and insuring the comparability of the weights via the latin square, the researcher may be able to increase the precision with which he can measure the effect of different window displays, on per customer apple sales.

The precision increasing techniques (pairing and the latin square design), that have been illustrated in this section, are concerned with insuring the comparability of experimental units prior to the administration of treatments. A statistical technique known as covariance analysis can also be used as a precision increasing device. In contrast to matching, covariance analysis²⁵ is based on supplementary observations that are not available at the time of the experiment. For example, matching could be used to group the experimental units by day-of-week or time-of-day, while covariance analysis might be used to adjust for the effect of variations from unit to unit in the average price of substitutes for apples. To illustrate a bit further, suppose we take 10 pairs of stores matched with respect to the day of the week on which measurements are to be taken. Ten stores will have a window display and 10 will have no window display. The per customer apple sales, observed during the experiment, will vary from store to store within both groups of stores. The greater the degree of variation within each of the two groups of stores, the greater will be the uncertainty associated with the results of the experiment.

²⁵ The dichotomy between experimental and observational designs, while pedagogically useful, is not quite accurate. An experiment which is combined with covariance analysis utilizes both types of designs.

Can this store-to-store variation in per customer apple sales, within the group of stores with window displays and within the group without displays, be reduced (thereby increasing the precision of the experiment)? Suppose we observe that stores in the experimental group, which tend to have high per customer sales, also have a higher average price for apple substitutes compared to apple prices, than stores which had lower per customer sales (that is, part of the store-to-store variation within experimental groups was positively associated with variations in the relative price of substitutes). Given this information, one can then adjust per customer apple sales in the experimental units for differences in the relative price of substitutes.²⁶

Testing the Effects of Two or More Experimental Variables

Thus far, the discussion has concentrated on examples of experimental design in which the effect of only one variable at a time was measured. One of the more common misconceptions with respect to experimental design is that only one variable at a time can be measured. In fact, there are numerous designs which measure the effects of more than one factor at a time. While a detailed discussion is beyond the scope of this book, there are a number of excellent texts devoted to the subject.²⁷ An illustration of one type of design capable of measuring the effects of more than one variable at a time is presented in the following paragraphs.

Suppose the management of a food chain wanted to test the effect of a new method of display for frozen food and a new type of window display. The effect of both of these changes could be evaluated in the context of the same experiment. For example, one could take one hundred stores and randomly divide them into four groups of 25 stores each, in the following fashion:

		<i>Window Display</i>		
		Standard	New	
<i>Frozen Food Display</i>	Standard	25A	25B	50
	New	25C	25D	50
		50	50	

²⁶ A more detailed exposition of the logic underlying covariance analysis is presented by Massy in the following paper.

²⁷ See Cox, *op. cit.*, Cochran and Cox, *op. cit.*, and Federer, *op. cit.*

This example is known as 2 x 2 factorial design, as there are two factors, window display and frozen food display, each of which is evaluated at two levels. In conducting the experiment:

- A. Twenty-five stores would have both the standard window display and ~~new~~ frozen food display.
- B. Twenty-five stores would have the new window display and the standard frozen food display.
- C. ~~Twenty-five stores~~ would have the standard window display and the new frozen food display.
- D. Twenty-five stores would have both the new window display and frozen food display.

A comparison of the performance (for example, per customer sales) of the 50 stores in groups A and B versus the 50 in C and D gives an evaluation of the effect of the new frozen food display. A comparison of the 50 stores in groups A and C versus the 50 in B and D gives an evaluation of the effect of the new window display. In addition, one has evidence as to whether or not the effect of introducing the new frozen food display is greater with the standard or the new window displays, or alternatively, whether the effect of introducing the new window display works better with the standard or the new frozen food display.²⁸

There is a fair amount of flexibility with respect to the types of designs that can be used. For example, there are designs that can evaluate the effects of more than two variables at once. In addition, more than two treatments can be studied for a given variable (for example, a design could be developed to choose between several different types of window displays, not just two).

In practice, the actual design of an experiment requires considerably more artistic skill than this discussion implies. Although the general principles set forth in our discussion of experimentation have a rather wide range of applicability, each specific problem has enough idiosyncratic characteristics of importance to make experience, as well as substantive knowledge, an essential asset for either the evaluation of or the conduct of research.

THE MEASUREMENT OF CAUSAL RELATIONSHIPS: OBSERVATIONAL STUDIES

There are many situations in which experimentation is not feasible or the most economical procedure. For example, one of the determinants of demand for Cadillac cars is family income. It is impossible to assign people to different income levels to study the effect of income on consumption of Cadillacs. Instead, a record is made of Cadillac ownership and income by family and from it an estimate of the relationship can be computed.

²⁸ The term "interaction" is used to refer to this situation (that is, interaction between two factors means that the effect due to one of them, for example, window displays, depends on the level of the other, for example, frozen food displays).

Whatever the form of a study, if it is to provide for the measurement of a causal relationship, it must provide a number of safeguards against unwarranted conclusions.

Nonexperimental studies cannot provide safeguards as adequate as those given by random assignment of subjects to experimental and control groups, direct manipulation of the experimental variable, and control over some of the extraneous variables that might operate during the course of the experiment. What substitute safeguards are available? For direct manipulation of the experimental variable, the investigator may substitute one or more of several lines of evidence: comparison of people who have been exposed to contrasting experiences, attempts to determine the time order of variables that are associated, examination of the relationship between variables in terms of the pattern of relationships that might be anticipated if one or the other were the causal factor. For assignment of subjects to experimental and control groups, the investigator may substitute evidence which provides a basis for inferring that [for measuring the extent to which] groups of people who have undergone contrasting experiences were or were not similar before those experiences; or he may select from his total group subsamples matched in terms of certain characteristics but with contrasting experiences; or he may restrict his sample to persons with certain characteristics. For direct control over extraneous variables, either past or contemporaneous, he substitutes the gathering of data on other characteristics or experiences of his subjects which he believes may be relevant to position on the dependent variable, and makes use of these data in his analysis.²⁹

Substitutes for Direct Manipulation of the Assumed Causal Variable

The comparison of groups exposed to contrasting experiences is probably the most common device used in lieu of direct manipulation in experimentation. Its usefulness in isolating the effect of the variable under study varies considerably from problem to problem. For example, a survey was made by a large grocery product manufacturer in order to determine the effectiveness of an advertising campaign in one city. The objective of the campaign was to change consumer attitudes toward a product with respect to its being fattening. It was found that the percentage of people who thought the product was fattening was lower among people who said they had seen the ad (56%) than among those who said they had not seen it (63%). However, it is rather difficult to place any confidence in these results because the members of the "saw" and "not-saw" groups were self-selected. It could be that people who remembered the ads tended to be people who agreed with the message in the first place so that the difference between the 63% and 56% was due to differences between the previous attitudes of the respondents in each group rather than to the effect of the ad.

This is not to say that the usefulness of matching groups with contrasting experiences is always limited by the effect of self-selection.

²⁹ Selltitz, Jahoda, Deutsch, and Cook, *op. cit.*, pp. 127-28. ✓

Alfred Politz has reported the results of a study concerned with automobile consumers' notion of pick up:

Among motorists who report that it is somewhat difficult to push and keep their accelerators down, 26% give their cars credit for good pickup. Among motorists who report that it is easy to push and keep their accelerators down, 61% give their cars credit for good pickup. This means one of two things: (1) either a soft accelerator spring contributes substantially to the motorist's estimate of his car's acceleration, or (2) most automobiles equipped with a soft spring also have a high actual acceleration rate.

The problem is cornered but not yet solved. We can identify the right alternative through engineering data. And it turns out that among car models with high acceleration, stiff and soft accelerator springs occur in approximately the same proportion as they do among car models with low acceleration. Hence, the second alternative must be wrong.³⁰

Two assumptions need to be met in this study if the results are to approximate those of an experiment: (1) that soft and stiff accelerator springs occur in the same proportion among cars of high and low acceleration and (2) that the selection of a given make is independent of the strength of its accelerator spring.

With no before measurement the researcher lacks data with which to verify the fact that the groups were comparable. However, in a situation such as the Politz example, it would seem unlikely that self-selection or some other form of systematic bias would have occurred.

Another substitute for direct manipulation is evidence as to the time order of variables. This is usually obtained in cross-sectional studies by asking respondents about time relationships, or by gathering evidence based on studies extended over time (time series analysis) or by some combination of the two (panel studies).

Deutsch and Collins,³¹ in a study of the attitudes of white tenants toward Negroes in public housing projects, were concerned about the effect on attitudes of living in an integrated versus a segregated project.

If, once the results were in, they found that whites in integrated projects had more favorable attitudes toward Negroes than those in segregated projects, the results would probably be held to be inconclusive due to the possibility of self-selection. Whites who had favorable attitudes might be more apt to choose integrated projects. In order to get information on past attitudes, respondents were asked such questions as: Can you remember what you thought Negro people were like before you moved into the project? How much have your ideas about Negro people changed since you have lived in the project? In effect, Deutsch and Collins were trying to get a before-after measurement as one might do in an experiment.

³⁰ Alfred Politz, "Science and Truth in Marketing Research," *Harvard Business Review*, Vol. XXXV (1957), pp. 119-20.

³¹ M. Deutsch and M. E. Collins, *Interracial Housing: A Psychological Evaluation of a Social Experiment*, University of Minnesota, 1951.

Another way to measure changes over time would be to take repeated surveys (cross-sectional analysis), each one with a new group of respondents, such as the public opinion polls conducted by Gallup and Roper, the census of population, and *Life's* survey of consumer expenditures.

An alternative research design, with an objective similar to that involved in asking people about changes in their behavior over time or taking repeated surveys with different respondents, is the repeated interviewing of the same group of respondents over time. These panel studies usually consist of having essentially the same group of respondents fill out periodic reports. The Market Research Corporation of America has a national panel of about 10,000 families that keep weekly diaries on a wide range of frequently purchased food and household items. A. C. Nielson Company has a panel of radio and TV listeners. Consumer panels are also used for such purposes as product testing and magazine readership.

When thought of as an approximation to an experiment, a panel can be pictured as having the following design:

	<i>Experimental Group</i>
Before measurement.....	Yes (X_1)
Second measurement.....	Yes (X_2)
First experimental variable.....	Yes
Third measurement.....	Yes (X_3)
Second experimental variable.....	Yes
Fourth measurement.....	Yes (X_4)
Fifth measurement.....	Yes (X_5)
Third experimental variable, etc.....	Yes

Panel studies can be differentiated from other types of observational studies in several ways:³²

1. With surveys taken over time, but with a new group of respondents each time (cross-sectional surveys) one can measure a brand's share of market over time. In a panel study one can also measure the amount of shifting activity between various brands over time.
2. In attempting to evaluate the effects of a change in a firm's promotional mix, panel data not only permit the contrast of exposed and unexposed respondents but also allow the researcher to contrast their brand switching behavior.
3. Repeated interviews with the same respondent increase one's chances of getting more information on that individual's behavior than would be possible with cross-sectional surveys.
4. By providing a record of behavior over time based on the submission of periodic questionnaires panel data decrease the degree of reliance that is placed on the respondent's memory. For example, panel data might provide a better basis for classifying respondents as regular, infrequent, or

³² This summary of a panel's advantages and disadvantages is paraphrased from Zeisel, *op. cit.*, pp. 217-19. For a more detailed discussion of the use of panel data see Zeisel, *ibid.*, pp. 215-54.

- never users of a brand during the past year than would a cross-sectional study, which asked the respondent into which of these categories she fell.
5. In cases where the researcher is interested in measuring change over time (for example, the change in market share from one period to the next) the statistical precision of a panel is often greater than that of a cross-sectional study of comparable size and structure.

The main disadvantages that differentiate panel studies from other types of observational investigations are the degree of mortality and the possibility of re-interview bias. In a panel study, the researcher is asking a respondent for continuous cooperation as opposed to a one-time interview. The refusal rate is, therefore, apt to be higher in panel studies. Panel studies, by definition, involve the re-interviewing of respondents. Individuals who are willing to cooperate may behave differently than the general population, or than people who are willing to cooperate with a one time interview. In addition, the very act of repeatedly interviewing a respondent may lead to changes in behavior.

Substitutes for Random Assignment of Subjects to Experimental and Control Groups

In an experiment the investigator can use matching to help assure the comparability of experimental and control groups. In the case of observational research, the investigator has three principal substitutes for matching: (1) evidence as to the initial comparability of the groups. (2) comparison of matched subgroups, and (3) restricting the sample.

In the case of the Politz study, there is the possibility that people might think a soft accelerator means high acceleration, and therefore pick a car with a soft accelerator. This possibility could be researched to see if self-selection was actually occurring. Such a study would provide information which would bear on the probable initial comparability of the two groups (that is, owners of cars with soft and stiff accelerators).

Studies of advertising effectiveness might ask a respondent what brand of product he was using prior to seeing an ad in order to determine the extent of possible self-selection, and thereby make a judgment as to the initial comparability of exposed and unexposed respondents with respect to previous consumption.

Katz and Lazarsfeld skillfully describe the technique of matching subgroups in their book, *Personal Influence*.³³ In their analysis of the determinants of the flow of influence in fashion, they found that fashion leadership declined as a woman progressed through her life cycle (for example, girls were more apt to be fashion leaders than wives with large families). They also found that interest in fashions declined with life cycle, and in addition, that fashion leadership was strongly related to interest. These results raised a question as to whether the relation be-

³³ Elihu Katz and Paul Lazarsfeld, *Personal Influence* (Glencoe, Ill.: Free Press of Glencoe, Inc., 1955), pp. 247-54.

tween leadership and life cycle was due to the fact that interest declined with life cycle or due to something else such as a decline in contact with people to be influenced. In order to shed light as to which of these factors was at work, they looked at the subgroups of interested women at all life cycle stages. They reasoned that if interest accounted for the decline in fashion leadership with life cycle, then the fraction of leaders in each high interest group should remain constant regardless of life cycle. If, on the other hand, fashion leadership declined as life cycle progressed among the high interest group, this would suggest that some other factor was at work. In other words, they used the device of dividing women at each stage of the life cycle into subgroups of high and low interest (in fashion). Then each of these subgroups were compared over all life cycle levels.

In a study primarily concerned with the ability of psychological factors (as operationalized in terms of the Edward's Personal Preference Schedule) to predict the brand choice of Ford and Chevrolet car owners, Franklin B. Evans³⁴ used the technique of sample restriction as a device to help insure that any differences that were observed in the psychological scores of Ford and Chevrolet owners were associated with personality attributes and not associated with some other underlying differences in say, social class or sex.

Evans restricted his sample in three ways:

1. He interviewed only the residents of a particular suburb, Park Forest, in an attempt to get a relatively homogeneous group with respect to demographic characteristics.
2. He included only Ford and Chevrolet owners of 1955-58 models in order to minimize the effects of the style cycle of these brands.
3. He interviewed only male owners having one car.

Though this restriction limits his ability to statistically generalize the results, it nonetheless serves as a device which leads to a greater degree of sensitivity with respect to discriminating between Ford and Chevrolet owners on the basis of their personality characteristics, than would be possible if the aforementioned characteristics had been free to vary.

THE LIMITATIONS OF EXPERIMENTATION³⁵

The limitations discussed in the quotation to follow are limitations of tryouts, generally, and the first four, at least, are limitations of non-experimental as well as experimental research. Some of them apply to tryouts as well.

1. *Experiments typically must be limited to measuring short-term response, but long-term response is often relevant in marketing problems.* It is no acci-

³⁴ "Psychological and Objective Factors in the Prediction of Brand Choice: Ford versus Chevrolet," *Journal of Business*, Vol. XXXII, No. 4 (October, 1959), pp. 340-69. See pp. 307-8 in this book.

³⁵ The following discussion with respect to the limitations of experimentation consists principally of a quotation from Howard and Roberts, *op. cit.*, pp. 16-19.

dent that the best and most extensive marketing experimentation thus far has been concerned with direct mail advertising and with store merchandising of *perishable* food products where the immediate sales response is of paramount interest. The problem of carry-over effects may be much more serious in other areas.³⁶ One experiment succeeded in measuring an improvement in car sales for a response period of three months but left legitimate doubt as to how much of the improvement may have represented borrowing from future sales. Other experiments have demonstrated small or negligible immediate sales increases but have provided no answer as to whether the sales action might have succeeded if it had been continued for a longer period. Many product innovations require changes in buyer habits and catch only after initial disappointments. Thus even an experiment that accurately measures short-term response may reduce so small a part of the uncertainty surrounding a decision that the experiment will be of limited usefulness.

2. *It is usually "expensive" to measure accurately actual sales in individual experimental units.* If families are the experimental units, either repeated interviews or consumer diaries are typically required. These are expensive and entail the risk that respondents will be "conditioned" by the process of measurement. If small geographic areas such as sales territories are used, accurate sales figures often can be obtained only with substantial added expense because accounting systems usually furnish delivery but not sales information. (Sometimes the records of wholesalers and distributors are helpful.) Special record-keeping devices must be established. When retail stores are the experimental units, measurement is easier, but even here special store audits or successive shelf counts may be needed. Even direct mail advertising experiments, where measurement is easiest, often entail special costs in tracking sales to individuals on the mailing list. (Some consumer panels can be used for experimental purposes.) "Expense" of measurement is, of course, meaningful for decision purposes only when contrasted with potential benefits; but it still must be said the expense of getting accurate sales measurements is often surprisingly large.

3. *The variability of sales—whether between families, dealers, geographical areas, or time periods—is often large by comparison with hoped-for responses to marketing actions.* A common criticism of experimentation in social research or marketing is that little or nothing can be learned because there are so many uncontrolled variables. Randomization prevents these variables from exerting any biasing effect on the average, but all the tricks of experimental design may not suffice to reduce random variation to the point where important responses can be accurately estimated unless economically exorbitant sample sizes are used. The amount of uncontrolled variation will, of course, depend on the particular problem. Sometimes this variation is startling, as in sales data by stores or territories where even experienced marketing men, who are fond of stressing individual differences, find it hard to believe that the observed variations are real.

While uncontrolled variation is discouraging, it does not necessarily rule out experimentation, but does demand careful quantitative study in the design stage of any experiment. A rigorous analysis of whether the company should be willing to pay for a larger sample, settle for a smaller one, or do no experimenting at all depends not only on the extent of random variation but also on costs of error, costs of sampling and available information about the probable effectiveness of the stimuli. Advance calculations of the effect of uncontrolled

³⁶ For a detailed discussion of one technique used for estimating the effects of carry-over see R. J. Jessen's article on switch-over designs in the section on field experiments.

variation always should be attempted, even if they rest in part on guesswork. Experimental costs should be estimated correctly. For example, an experiment should be assessed only incremental cost such as cost of split-run mailings of catalogs and tracing of sales responses; the basic media cost, on the other hand, would be incurred even if an experiment had not been run. This simple point seems often to have been misunderstood in practice.

4. *It is difficult to prevent "contamination" of control units by test units.* Few marketing stimuli can really be directed only to specific units. Local advertising media spread out widely from their area of primary focus, stores have overlapping trading areas, and recipients of mail order catalogs talk with each other. Training experiments for sales personnel can rarely prevent transmission of information from test personnel to control personnel. All these illustrations are examples of "contamination": the experimental stimuli reach not only the test units, but via the test units they sometimes reach some of the control units as well, thus making the test-control comparison obscure.

The reader is reminded that these first four points are at least as relevant in nonexperimental as well as experimental investigations. The remaining limitations are usually but not necessarily more likely to be important in experiments than otherwise.

5. *It is often difficult except for very short periods to execute experimental designs properly when they require people in a marketing organization to behave differently than they would otherwise have done.* A store merchandising experiment for a private company . . . required changes in shelf allocation in supermarkets to be made and retained for several weeks. In spite of many prudent measures to assure compliance, store managers found ways to delay or to avoid making the intended changes. Difficulties of this kind probably occur more frequently than are reported. We have found few store experiments in which store managers were required to alter shelf-allocation or display for extensive periods. A number of the Cornell store experiments,³⁷ however, extended over six weeks to two months. In one study of pricing, . . . , the store manager was reimbursed for the losses that might result from charging below-market prices. Finally, in many of the store studies, "enumerators" to insure the experiment was being properly carried out were in the store all of the time or checked it frequently, e.g., twice each day.

Those dealers or territories used as "controls" may learn or be told this fact; if so, their opposition or resentment may then be a real barrier to execution of the design. We know of at least one proposed large scale experiment that was dropped for this reason. Thus "people" are reluctant to carry out experimental designs. It is interesting to see how this obstacle differs from a traditional obstacle to experimentation in social science that also centers on "people." It is commonly said that *people as subjects* (or better, objects) are reluctant to modify their behavior in the interest of an experiment. In marketing, and business generally, it is often possible to experiment on people without their knowledge that an experiment is taking place. But those who apply the marketing stimuli—those in the marketing organization—do know that an experiment is taking place, and as explained above, are often reluctant to participate as directed. To the extent that people as subjects are aware that they are participating in an experiment, the "traditional obstacle" to experimentation is, of course, real. Thus sales trainees may do better than a control

³⁷ This refers to a number of experiments, the best known of which are concerned with the in-store promotion of McIntosh apples, conducted under the direction of Professor Max E. Brunk, Department of Agricultural Economics Cornell University Agricultural Experiment Station, New York State College of Agriculture.

group because they are pleased to have been specially chosen for the program. Closely related is the so-called "placebo effect," well known in medical research, in which an inert treatment may improve the recovery rate above that in absence of treatment. To the extent that this possibility is real in a marketing experiment, it may be possible to use "placebo" group, in addition to control and test groups, preferably forming all groups randomly. The difficulty is that "inert" treatments are much harder to devise in marketing than in medicine.

6. *It is often difficult to make marketing experiments sufficiently realistic to be useful.* National advertising media often have to be used intact and thus cannot be adapted to experiments in local areas. The National Opinion Research Center conducted for the NCAA a series of television experiments to measure the effect of TV on local football attendance by actually "blacking-out" some areas but not others and measuring comparative attendance. But this is the only illustration we have been able to find in which network television was used experimentally by local areas. Aside from split run tests, we know of no analogous marketing experiments for printed media. Sales tests of advertising have often turned to less realistic situations such as door-to-door selling, use of handbills in the trading areas of supermarkets, point-of-sale advertising display, consumer jury tests, etc. Another problem in attaining realism stems from the fact that competitors may not respond to local experimental actions but then respond later to these actions if adopted as a general policy. At the other extreme competitors may be unusually energetic in "jamming" the experiment once they learn of it.

7. *Experiments raise "security" problems of a more serious nature than those associated with surveys.* An experiment, by its definition, reveals a proposed marketing action that might otherwise have been concealed or at least exposed more discreetly in a survey. We have found two examples of effective "counter-intelligence" by competitors who were able to learn the results of their competitor's new product test and to market the product themselves before the experimenter.

8. *The mortality of experimental units is relatively high in marketing experiments.* During the course of marketing experiments, stores have been hit by hurricanes, salesmen have quit, consumers have moved, records have been lost, etc. Thus, there may be "missing observations" at the end of the experiment. Missing observations represent a shrinkage of sample size. If they occur at random in a simple experimental design, they hardly are more serious than the reduced sample size would suggest. In more elaborate layouts even randomly occurring missing-observations, especially if numerous, make an exact statistical analysis of the results very costly. Almost all "fancy designs" that we have found were used in experiments of relatively short duration so that few observations were apparently lost. We expect, however, that the missing-observation problem will continue to encourage the use of relatively simple designs in long-duration experiments.

We have been primarily concerned with the determination of cause and effect relationships via experimentation. The following outline is presented in an attempt to facilitate the reader's analysis of the articles on experimentation which follow in a later section of this book:

A GUIDE FOR THE ANALYSIS OF EXPERIMENTS

- I. What is the relationship between the marketing decision under consideration and the research design used in the investigation?

- A. In terms of variables considered:
 - 1. What variables are apt to be important determinants of the decision?
 - 2. Of these which are included in the research design?
 - 3. What would you expect the effect to be of modifying the design so as to include more of the important determinants of the decision?
 - 4. If such a modification is desirable from the standpoint of the problem, is it feasible in terms of its cost and the complexity of the research design?
 - B. In terms of the physical context of the research:
 - 1. Was the experiment performed in the field or the laboratory?
 - a) If it was performed in the laboratory:
 - i) What were the laboratory conditions?
 - ii) How were these different from those under which the policies being evaluated will be implemented? (That is, is the motivation of the respondent apt to be the same in the laboratory as in the field?)
 - iii) In what way would you expect the observed results to change if some of these conditions were modified?
 - iv) Could the existing study be modified so as to bring the nature of the laboratory situation closer to that of the "real" world?
 - v) If the answers to the previous question was yes, why do you think these modifications were not made when the design was originally implemented?
 - b) If it was performed in the field: ask questions (a) i to (a) v.
 - 2. Was the group of experimental units (individuals, stores, cities, and so on) apt to reflect the characteristics of the population to be affected by the decision? If not:
 - a) What are the differences apt to be between the population being investigated and the one that is being acted upon by the decision maker?
 - b) If the population being investigated were modified so as to bring it closer to reflecting the characteristics of the population to be acted upon, in what way would you expect the results of the investigation to change?
 - c) What modifications would you make? Why?
- II. What is the nature of the experimental design under investigation?
- A. How were the control groups organized?
 - 1. Were before and after measurements taken?
 - a) Was there apt to be any interaction between the before measurement and the experimental variable? Between the before and after measurement?
 - b) If the answer to either of the questions on interaction is yes, then in what ways were the control groups organized to avoid these interactions?
 - c) Could some other procedures be used?
 - 2. Prior to the administration of the treatments, what was done to insure the comparability of the groups of units to be exposed to each treatment?
 - a) In what ways could these procedures be modified so as to improve one's basis for comparison?
 - b) What are the costs (money, time, and complexity) involved in the modifications? Do you think they are justified?

3. Was any supplementary data collected during the administration of the treatments in order to facilitate their comparison when the results were analyzed?
 - a) What other supplementary information might have been desirable to collect?
 - b) What would you do with the information if you had it?
 - c) How would you expect it to modify the results?
- B. Was randomization employed in the process of assigning treatments to experimental units? Why or why not, as the case may be?
- C. In what fashion was the experimental response measured? Would some alternative unit of measurement be apt to yield more clear-cut results?
- D. In cases where supplementary information was collected, what units of measurement was used with respect to each variable?
 1. Were these operational definitions appropriate for the problem at hand?
 2. How could they be modified so as to be more useful in terms of coming closer to a better measure of the variables under investigation in the experiment?
- III. What are the implications of the analysis of this experiment with respect to possible applications to other marketing problems or to other product categories?
 - A. What were the characteristics of the product and/or the problem under investigation?
 - B. In what way do these characteristics differ with respect to other products and/or problems?
 - C. How are these differences apt to affect the potential contribution of experimentation to the solution of the problem? Or, to put it another way, with respect to what types of problems and/or product categories is experimentation apt to generate its most fruitful results in terms of its contribution to the evaluation of alternative management policies?

The following chapter will concentrate on the use of statistical analysis as a basis for measuring the direction and magnitude of causal relationships.

SUGGESTED READINGS

- BRUNK, MAX E., and FEDERER, WALTER T. *Methods of Research in Marketing*. Paper No. 4, Department of Agricultural Economics, Cornell University Agricultural Experiment Station, New York State College of Agriculture. Ithaca, N.Y.: Cornell University, January, 1953.
- COCHRAN, WILLIAM G., and COX, GERTRUDE M. *Experimental Designs*. New York: John Wiley & Sons, Inc., 1957.
- COX, D. R. *Planning of Experiments*. New York: John Wiley & Sons, Inc., 1958.
- DOMINICK, BENNETT A., JR. *Research in Retail Merchandising of Farm Products—Appraisal of Methods and Annotated Bibliography*. Marketing Research Report No. 416, Market Development Research Division, Agricultural Marketing Service, U.S. Department of Agriculture. Washington, D.C.: U.S. Government Printing Office, 1960.
- FEDERER, WALTER T. *Experimental Design*. New York: The Macmillan Co., 1955.

KEMPTHORNE, OSCAR. *The Design and Analysis of Experiments*. New York: John Wiley & Sons, Inc., 1952.

SELLTIZ, CLAIRE; JAHODA, MARIE; DEUTSCH, MORTON; and COOK, STEUART. *Research Methods in Social Relations*. rev. ed. New York: Henry Holt and Co., Inc., 1960.

ZEISEL, HANS. *Say It with Figures*. rev. 4th ed. New York: Harper & Bros., 1957.

Statistical Analysis of Relations between Variables

WILLIAM F. MASSY

THE PREVIOUS ARTICLE DEALT WITH THE STRATEGY OF RESEARCH DESIGN. The design of causal studies was discussed in depth, because the business firm needs cause and effect data as an input to its decision-making process.

Knowledge of a cause and effect link implies that we understand a relationship between two or more variables. We must search for independent or explanatory variables that (1) account for a substantial fraction of the behavior of the dependent variable under study, and (2) appear reasonable in terms of our subjective understanding of the problem. This chapter is devoted to the study of statistical techniques that can help to determine the form and extent of the relationship between a dependent and one or more independent variables.

Introduction to the Analysis of Relationships. The procedures to be discussed below are designed to provide measures of *statistical association*, that is, the extent to which two or more variables tend to change together. Association alone, however, can never be sufficient to establish a causal link between different variables, since statistical measures of association are very sensitive to the effects of extraneous variables. An analyst who obtains a high correlation (a measure of statistical association to be discussed below) between the birth rate in India and the United States gross national product, for example, had best think twice before concluding that one *causes* the other. Statistical association is only half the story; few reasonable men would agree that there is a causal link between the two. In this case, both births and GNP are growing through time as the result of their own (independent) causal mechanisms; since both are highly correlated with time, they are correlated with each other. This effect is known as *spurious correlation*, which might better be called *spurious causal association*. We emphasize that there is nothing wrong with the correlation measure. It shows a real statistical association,

but the conclusion of causality is spurious because of the extraneous factor, time.

The techniques to be introduced below can be used with either experimental or observational data. The reader will recall that the major purpose of an experimental design was to control the effects of extraneous variables in order to prevent them from cluttering up the results and obscuring the relationship being studied. We have seen that the validity of causal interpretations based on statistical association depend upon the effects of variables that have not been specifically included in the analysis. Statistical association is often a reliable indicator of causality in experimental designs, because excluded factors are usually controlled. For observational designs, on the other hand, real art is required in order to assess the effects of uncontrolled (and usually unmeasured) variables. Much of the discussion presented below will be aimed at developing criteria for deciding whether the effects of an excluded variable are likely to spoil the causal interpretation of a statistical analysis.

Many of these criteria can only be evaluated in terms of the analyst's subjective judgment: there are few statistical tests which can provide "scientific" assurance that the data's meaning is really what it appears. Unfortunately, many people believe that data speak for themselves. They do not, and any analyst who looks *only* to his tables and charts is in for some rude surprises. In the sections that follow we will see that the same set of data can be interpreted in many different ways. Only by thinking over what he already knows about the particular problem in question can the analyst decide which of the alternative conclusions is the most reasonable.

It does not make sense to throw away knowledge obtained from previous experience when we begin the analysis of a particular set of statistical data. Besides its critical role in the interpretation of causal relations, prior knowledge can be put to work in the following ways:

1. We can choose the *statistical procedure* which best fits our problem and the type of data which is available.
2. We can choose a *structural model* to provide a framework in which to apply the statistical procedures and interpret their results.
3. We can assess the *amount of information* contributed by our statistical analysis in terms of our needs and what we already know.

Problems (2) and (3) will be discussed briefly here; (1) will be an independent part of each of the following sections.

Models. A model can be viewed as an hypothesis about the way the world operates. In the most general sense, it is a *collection of statements about the way in which certain variables are causally related to one another.*

All of us make use of models in carrying on our everyday lives. Business decisions are no exception to this rule. A marketing executive might say to himself:

If I cut my price to customer X in Kansas City, my largest competitor will cut his everywhere. I believe this for two reasons: (1) the competitor's general pricing policy, as publicly stated on numerous occasions, implies this kind of retaliation; and (2) when we tried to shade a price in Cincinnati a few months ago, he *did* cut across the board. A national price cut on this product would be disastrous; therefore, I will hold the line in Kansas City.

The foregoing is an example of a simple model of competitive behavior. The executive probably has a high degree of confidence about the truth of his hypothesis, since it is based on both a priori and empirical evidence: the competitor has said he would cut price if challenged in any one market, and he did. The forecast of the competitor's behavior, which was based on this model, is highly believable.

If, on the other hand, our hypothetical executive knew only about his competitor's reaction to the price cut in Cincinnati, it would be very dangerous to infer that the action was part of a consistent long-run policy and would be repeated under similar circumstances; many other equally plausible reasons are possible. By supplementing this observation with subjective knowledge about the competitor's general policy, acquired gradually over a period of time, the behavior actually observed takes on new meaning. We will see that many statistical techniques derive their power from the ability of the user to specify the kind of relationship he expects to obtain.

Models must be rich enough to take account of the relevant and important variables. Nevertheless, complication is not desirable per se. The temperament of the secretary to the president of the competing company would not normally be included in the model given above (although that of her boss might). One of a model's advantages is that it is an *abstraction* from the real process being described. Only the variables that are relevant and important are included, allowing attention to be focused upon them and not diverted along blind alleys. This view is compatible with the principle of *parsimony of variables*, from the philosophy of science: the simplest model that can adequately describe the phenomenon under study should be adopted. On the other hand, too much abstraction reduces the model's usefulness in the solution of real problems.

Verbal models are often inefficient tools for the interpretation of empirical data. If the data are quantitative, it is usually best to describe the model in terms of mathematical equations. The mathematical model implies no more than the word model and can be constructed using similar information. Its advantage is that it is unambiguous and lends itself to algebraic manipulation and statistical analysis. The "relationships" discussed in this paper are usually posed in mathematical terms. They are nothing more nor less than simple models of the behavior under study.

Information. It is easy to see that an analysis of empirical data will contribute some information (however small in amount) to the problem

under study. The amount of information so contributed can be measured in many kinds of problems. In turn, this provides the basis for appraising the *efficiency* of alternative statistical procedures.

The concept of information is closely related to a well-known measure of statistical dispersion, the *variance* (the square of the standard deviation). Denoted by the symbol σ^2 , the variance of any statistical quantity is defined as the average of the squared deviations of that quantity from its mean or expected value. While the variance may serve merely as a summary of the degree of dispersion present in a particular set of known numbers, it usually is used as a measure of the amount of random fluctuation associated with the estimate of a statistical parameter, of the kinds discussed in later sections. The theory of random errors and the calculation of variances and standard deviations are covered in almost all basic statistics texts.¹

A statistical estimate of the value of any quantity is subject to errors of estimation; their source will be discussed in connection with the various statistical procedures covered below. For now, it is enough to understand that the *precision* of any estimate is a function of the likely size of these errors. Providing that no bias is present, precision is measured by the variance of the estimate. The smaller the variance, the more precise is the estimate on average.

Bias can occur for one of two reasons:

1. The methods by which the original data were collected were incorrect. If the data are obtained by sample survey, for example, bias may be introduced by: (a) improper procedure in selecting the sample, leading to a sample that was not representative of the underlying population (for example, through failure to follow up on the nonrespondents); or (b) errors in the measuring process itself, such as faulty interviewing methods or questionnaire design, leading to answers that do not mean what was intended (for example, through inadvertently asking "leading" questions). These problems are extremely important for marketing analysts, but they are beyond the scope of this paper.

2. The statistical procedures used to analyze the data were improperly specified. Later in this paper we shall see that some statistical techniques will, if improperly used, yield biased estimates of the effects of particular variables, even though the original data are unbiased.

The ideas about information developed below are based on the assumption that both the data and the statistical procedures utilized are unbiased. The latter assumption will be relaxed in the subsequent sections, but we will continue to assume that our data are "good" throughout the paper.

While the variance of an estimate may be used to test hypotheses or to establish confidence intervals according to the standard techniques of statistical inference,² we will consider only its role as a measure of the

¹ Cf., William A. Spurr, Lester S. Kellogg, and John H. Smith, *Business and Economic Statistics* (rev. ed.; Homewood, Ill.: Richard D. Irwin, Inc., 1961), chaps. ix-xiii.

² *Ibid.*, chaps. xiii and xiv.

amount of information which the estimate contributes to our knowledge about the true state of the world.

For unbiased estimates with normally distributed errors, for which the variance is known, the amount of information contained is defined as the reciprocal of the variance.³ If we are trying to estimate the average income of families in the United States, for example, we would define the amount of information contained in a given sample of families as

$$I_s = \frac{1}{\sigma_{\bar{x}}^2}$$

where $\bar{X}_s$ is the average income for families in the sample (s), and I_s is the amount of information in (s). If the sample median (the income of the middle family, where all families have been listed in order of income level) had been used as an estimate of the average income of the population, the amount of information available would have been smaller: other things being equal it can be shown that the variance of the sample median is larger than that of the sample mean. The sample median is said to be a less *efficient* estimator of the population average than is the sample mean.⁴

It is obvious that if two statistical procedures measure the same thing and cost about the same, we would do better to use the one that was more efficient, that is, the one that contained the most information. When we speak of reducing the variance of an estimate, increasing the amount of information contained in it, or improving its efficiency, we will be talking about the same thing. We will want to use statistical procedures that are as efficient as possible, given the level of accuracy required in the final results and the amount of money we are able to spend.

The simplest way to learn about relations between variables is to form the data into an array. If done graphically, the array is called a *scatter diagram*, or, if graphical analysis is not desired, the data can be arranged in a *cross-classification table*. This technique will be discussed in the next section.

While arrays offer a quick and relatively easy method for evaluating gross relationships in data, a more objective form of analysis is required if subtle and partially hidden relationships are to be uncovered. A number of statistical procedures have been developed for measuring various types of relationships which may be found in empirical data. The most important ones will be discussed below: (1) correlation, (2) regression,

³ Robert Schlaifer, *Probability and Statistics for Business Decisions* (New York: McGraw-Hill Book Co., Inc., 1959), p. 443.

⁴ For large samples the sample median provides unbiased estimates of the population mean with 64 percent of the efficiency of the sample mean: R. L. Anderson and T. A. Bancroft, *Statistical Theory in Research* (New York, McGraw-Hill Book Co., Inc., 1952), p. 95.

(3) discriminant analysis, and (4) factor analysis. While other, more powerful, techniques are available,⁵ one of these four will often be adequate for handling marketing data.

CROSS-CLASSIFICATION

Arrays are designed to allow the visual interpretation of the relation between two or more variables. The idea can be illustrated in terms of

TABLE 1
Warm BRAND CANINE COATS
Yearly Sales by Metropolitan Area and
Effective Buying Income and Average Winter Temperature

Area*	Warm Brand Sales (Per Licensed Dog)† (Units × 10 ⁻³)	Effective Buying Income‡ per Family (\$ × 10 ³)	Average January Temperature§ (°F)
Atlanta.....	2.9	7.15	44.0
Baltimore.....	4.5	7.31	35.5
Chattanooga.....	1.6	5.54	42.5
Chicago.....	5.1	8.17	25.7
Cincinnati.....	2.6	6.74	32.6
Dallas.....	3.5	6.79	45.8
Detroit.....	4.5 ^a	8.10	25.5
Fort Lauderdale.....	1.9	5.79	68.0
Houston.....	4.1	6.86	54.2
Macon.....	1.1	5.76	46.8
Miami.....	3.4	7.13	68.0
Milwaukee.....	5.3	7.69	20.6
Minneapolis-St. Paul.....	3.7	7.13	13.1
Mobile.....	2.6	5.92	52.8
New Orleans.....	4.2	6.46	53.5
New York.....	5.5	8.36	32.1
Oklahoma City.....	1.3	6.23	37.6
Philadelphia.....	5.4	7.97	34.4

* Standard metropolitan areas as defined by the U.S. Department of Commerce.

† Hypothetical.

‡ *Survey of Buying Power* (Sales Management, 1961).

§ *Rand McNally Commercial Atlas and Marketing Guide*, 1958.

some hypothetical data on an "apparel" firm's sales. Let us assume that the company has prepared data on its sale of dog sweaters for a sample of 18 metropolitan areas in the United States. These data, together with that for consumers' effective buying income and an index of minimum winter temperatures in each of the areas are presented in Table 1.

⁵ In particular, regression yields biased parameter estimates where the equation being fitted is part of a simultaneous system of relations (predictions are unbiased as long as the structure remains constant, however). Maximum likelihood procedures for estimating parameters in certain kinds of simultaneous equation systems have been developed, but are beyond the scope of this book. See William C. Hood and Tjalling C. Koopmans (eds.), *Studies in Econometric Method* (New York: John Wiley & Sons, Inc., 1953), or any of the econometrics textbooks. A considerably easier "two stage least squares" method appears to be very promising as well. See Henry Thiel, *Economic Forecasts and Policy* (Amsterdam: North Holland Publishing Co., 1958).

We begin by looking at sales and income. While Table 1 presents the data for inspection, it is not an *array* because no effort has been made to throw the assumed relationship into prominence. In contrast, Figure 1 gives the *scatter* of sales values upon income. Each point on the diagram refers to a particular metropolitan area, with its Y and X coordinates being that area's sales and income values, respectively. High income areas tend to have high sales, and vice versa, although the relationship is not perfect.

Two other kinds of arrays can be based upon Table 1. Both are cross-classification tables but each focuses on a different attribute of the data. Table 2-A is a tabular summary of the scatter diagram. The values of

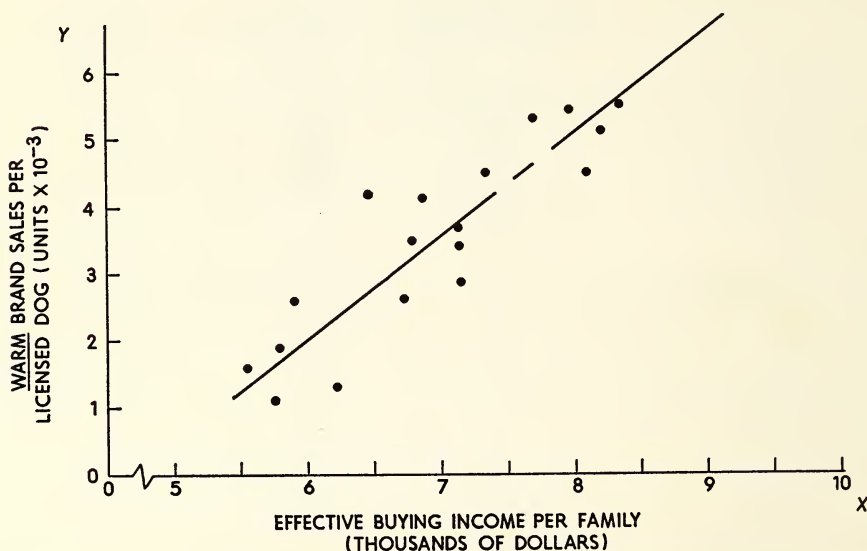


FIGURE 1. Scatter of sales on income for Warm brand canine coats. Based on the data in Table 1.

sales and income are divided into ranges, and the number of points falling into each cell counted. The resulting table, often called an *enumeration table*, thus gives the number of geographic areas having the specified magnitudes of sales and income. Table 2-B presents another kind of summary: the average value of sales for all the areas having the specified levels of income. It focuses attention directly upon the assumed dependent variable, sales, and is sometimes called an *attributes table*.

Which of the various arrays is best can only be decided within the context of a particular problem. The scatter diagram would probably be preferred in the present case because it gives the most direct information about the relationship under study, that is, the covariation of sales and in-

TABLE 2
CROSS-CLASSIFICATION OF SALES AND INCOME FOR
WARM BRAND CANINE COATS*

A. Number of Areas Having Specified Values
for Sales and Income

Income (\$1,000)

Sales
($\times 10^{-3}$)

$Y \backslash X$	5.5-5.99	6.0-6.49	6.5-6.99	7.0-7.49	7.5-7.99	8.0-8.49	Row Totals
5-5.9	0	0	0	0	2	2	4
4-4.9	0	1	1	1	0	1	4
3-3.9	0	0	1	2	0	0	3
2-2.9	1	0	1	1	0	0	3
1-1.9	3	1	0	0	0	0	4
Column totals	4	2	3	4	2	3	18

B. Average Sales for Areas Having Specified
Values for Income

Income (\$1,000)

$\bar{Y} \backslash X$	5.5-5.99	6.0-6.49	6.5-6.99	7.0-7.49	7.5-7.99	8.0-8.49	Grand Average
Average sales ($\times 10^{-3}$)	1.8	2.9	3.4	3.6	5.3	5.0	3.5

* Based on the scatter diagram in Figure 1.

come. If the number of areas in the sample were very large, however, the scatter diagram would be inconvenient relative to the two tabular forms. If more than two variables must be considered at the same time, arrays of the form of Table 2-B are usually best; the following variations can handle up to four variables at once:

Three Variables

(X, Y, and Z)

$Z \backslash X$	X_1	X_2	X_3	X_4
Z_1	$\bar{Y}_{11}$	$\bar{Y}_{12}$	$\bar{Y}_{13}$	$\bar{Y}_{14}$
Z_2	$\bar{Y}_{21}$	$\bar{Y}_{22}$	$\bar{Y}_{23}$	$\bar{Y}_{24}$
Z_3	$\bar{Y}_{31}$	$\bar{Y}_{32}$	$\bar{Y}_{33}$	$\bar{Y}_{34}$

Four Variables

(W, X, Y, Z)

Z \ X	X ₁			X ₂			X ₃		
	W ₁	W ₂	W ₃	W ₁	W ₂	W ₃	W ₁	W ₂	W ₃
Z ₁	$\bar{Y}_{111}$	$\bar{Y}_{112}$	$\bar{Y}_{113}$	$\bar{Y}_{121}$	$\bar{Y}_{122}$	$\bar{Y}_{123}$	$\bar{Y}_{131}$	$\bar{Y}_{132}$	$\bar{Y}_{133}$
Z ₂	$\bar{Y}_{211}$	$\bar{Y}_{212}$	$\bar{Y}_{213}$	$\bar{Y}_{221}$	$\bar{Y}_{222}$	$\bar{Y}_{223}$	$\bar{Y}_{231}$	$\bar{Y}_{232}$	$\bar{Y}_{233}$
Z ₃	$\bar{Y}_{311}$	$\bar{Y}_{312}$	$\bar{Y}_{313}$	$\bar{Y}_{321}$	$\bar{Y}_{322}$	$\bar{Y}_{323}$	$\bar{Y}_{331}$	$\bar{Y}_{332}$	$\bar{Y}_{333}$

While it is possible in principle to cross-classify any number of variables simultaneously, the method becomes unwieldy for more than four or five. The analytical procedures discussed in the following sections generally should be used for problems having larger dimensions.

Advantages. The use of arrays for evaluating relationships between variables in observed data offers the following advantages: (1) they are easy to understand and relatively quick to prepare; and (2) they require specification only of the variables to be included, rather than the form of the relation (as is the case with the analytical techniques discussed below).

Scatter diagrams and cross tabulations can be prepared by relatively unskilled personnel using conventional tabulating or punch-card equipment. While the time and cost of preparation may be substantial for large samples on a number of variables, arrays are usually cheaper to obtain than analytical results. Finally, arrays that are correctly specified and prepared can be evaluated effectively by people who have a knowledge of the problem area but who are untrained in statistical procedures. This consideration is so important that many market research analysts prepare arrays for presentation to management even where their own conclusions and recommendations are based upon the application of more complicated statistical techniques.

Less prior knowledge is required for the construction and interpretation of arrays than for analytical statistical procedures. In particular, the techniques to be discussed below all depend upon the *linearity* of the relationship under study. If the relationship is not linear, or if it cannot be made linear by a suitable transformation of variables, the statistical procedures may not yield valid results. The use of arrays avoids difficulties of this kind, since the form of the relationship can be inferred after the data has been arrayed rather than prior to the start of the analysis.

Specifying the Cross-Classification. Problems of specification can be illustrated within the context of our sales-income example. We will introduce a new hypothetical variable, type of area (that is, urban or rural). Suppose that for a constant level of income, sales are different for urban

than for rural areas. Let us further assume that the analyst is interested only in the effect of income on sales. He may then ask the question: "I am not interested in the effect of type of area on sales, but must I include this variable in my array in order to isolate the true effects of income?"

Alternative Specifications. There are three alternatives available to the analyst; each of them will be considered in this section:

1. Include urbanness in the cross-classification.
2. *Balance* the proportion of urban and rural areas in each income class, and *pool* the two sets of data.
3. Neglect type of area entirely, and *pool* the urban and rural data without balancing.

We shall see that alternative (1) is the best, while (2) sacrifices some information that is available in the data and (3) is likely to yield completely misleading results. (3) will be considered first.

Compare the figures in Tables 3-A and 3-C. The former shows that

TABLE 3
EFFECTS OF POOLING AND BALANCING ON CROSS-CLASSIFICATION OF
WARM BRAND CANINE COATS DATA

A. The True Array for Sales, Income, and Urbanness

Income (\$1,000)

	5.5- 5.99	6.0- 6.49	6.5- 6.99	7.0- 7.49	7.5- 7.99	8.0- 8.49	Grand Averages
Urban areas*	(4)† 1.8§	(2) 2.9	(3) 3.4	(4) 3.6	(2) 5.3	(3) 5.0	(18) 3.5
Rural areas†	(1) 0.2	(1) 0.3	(2) 0.4	(3) 0.5	(5) 0.7	(6) 0.8	(18) 0.6

B. Balanced Marginal Array for Sales and Income

*(Based on 50 Percent Urban and 50 Percent Rural
Areas in Each Income Class.)*

Balanced column averages	1.0	1.6	1.9	2.1	3.0	2.9	2.1
--------------------------	-----	-----	-----	-----	-----	-----	-----

C. Direct Marginal Array for Sales and Income

*(Based on the Raw Proportions of Urban and
Rural Areas in Each Income Class.)*

Direct column averages	1.5	2.0	2.2	2.3	2.0	2.2	2.1
------------------------	-----	-----	-----	-----	-----	-----	-----

* From Table 2-B.

† Hypothetical.

‡ Number of areas represented in the cell.

§ Average sales for all areas in the cell.

with one exception, sales tend to rise sharply with income for both urban and rural areas, but that sales in the rural markets are relatively small. The directly pooled data in Table 3-C are the same as would be obtained if the importance of the urban-rural distinction had never been recognized. They exhibit a considerably different picture than does Table 3-A: only for the lowest income class is a relationship with sales discernable. Changes in the composition of the sample have obscured the real effect of income upon sales. To anticipate the section on regression, we would say that the relationship exhibited by the directly pooled data is a *biased* estimate of the true relationship. The bias occurred because an excluded variable, type of area, was (1) an important determinant of sales; and (2) related to the level of income in the sample, in the sense that the higher income areas tended to be rural, and vice versa.⁶

Table 3-B shows the effect of balancing the proportion of urban and rural areas represented in each income cell. The balancing process begins with an evaluation of the frequency with which each value of the excluded variable occurs in the entire sample; in this case exactly 18 out of 36 areas are urban, and 18 rural, for a 50-50 split. Then the data in each of the income cells is pooled in such a way that the proportion of urban areas represented in each is the same as for the sample as a whole (in this case, 50 percent). Table 3-B exhibits the same general relationship of sales with income as was shown in Table 3-A. Given independence of effects, to be discussed later, it can be shown that bias in pooled data is eliminated by balancing out the contribution of the excluded variable, cell by cell.

In summary, *if any important variable is not included in the array the analyst must be sure that its total impact upon the reported value of the dependent variable is the same for all cells in the cross-classification. This equality of impact is achieved by balancing. It is essential if misleading conclusions are to be avoided.* (The data in Tables 1 and 2 were assumed to refer only to metropolitan areas, making the proportion of rural areas uniformly zero.)

Is Balancing Desirable? Should all relevant and measurable variables be included in the array or is cell by cell balancing of a variable's total impact upon the dependent variable a worthwhile short-cut? While working with a number of independent variables simultaneously is a difficult and time-consuming task, reducing the dimensions of an array through balancing is often undesirable for the following reasons:

1. Balancing eliminates bias, but does not utilize all of the sample data efficiently; that is, information is lost.
2. The validity of balancing depends upon independence of the effects of

⁶ In the language of regression, direct pooling throws the effects of urbanness into the error term. Condition (2), above, violates regression assumption 5 below (that the errors must be uncorrelated with the independent variables), which is sufficient to cause the estimate of the effect of income upon sales to be biased.

the (included) independent variable and the excluded variable being balanced out.

Each will be considered below.

1. Table 4 and Figure 2 illustrate the loss of efficiency due to excluding an important independent variable, even though balancing has been accomplished. The figure shows hypothetical sales values for three geographic areas in each of the four cross-classification cells presented in

TABLE 4
LOSS OF DISCRIMINATING POWER OF A CROSS-CLASSIFICATION OF
SALES AND INCOME DUE TO COMBINING DATA FOR URBAN AND
RURAL AREAS

A. The Full Array

	<i>Average Sales by Income Class*</i>		<i>Differ- ence A - B </i>	<i>Range</i>	<i>Ratio of Difference to Range</i>
	<i>A</i>	<i>B</i>			
Urban areas	(3) [†] 35 [‡]	(3) 40	5	12	0.42
Rural areas	(3) 15	(3) 20	5	12	0.42

B. The Balanced and Pooled Array

Pooled data	25	30	5	32	0.16
----------------	----	----	---	----	------

* New hypothetical data from Figure 2.

[†] Number of areas represented in the cell.

[‡] Average sales for all areas in the cell.

Table 4-A. Table 4-B gives the same data in pooled form. (It is also balanced, since there were three observations in each cell of Table 4-A.) Since we wish to evaluate the effect of income upon sales, our attention will be focused on the difference between the average sales in the two income classes. This difference is five units for both rural and urban areas separately and for the pooled data, which indicates that pooling the balanced data of Table 4-A did not introduce bias.

Efficiency has suffered, as can be seen by the ratios of the difference in average sales to the range of sales values for the cross-classified and pooled data. The range of a group of values is an estimator of the group's variance, which implies in turn that the range is related to the variance of the differences between the average sales figures for the two income groups. Therefore, we are justified in using the ratio of the difference to the range as a rough measure of the amount of information about the effects of income that has been extracted from the sample.

Look at the ratios for the cross-classified and pooled data. The former are much larger than the latter, which indicates that the variance of the

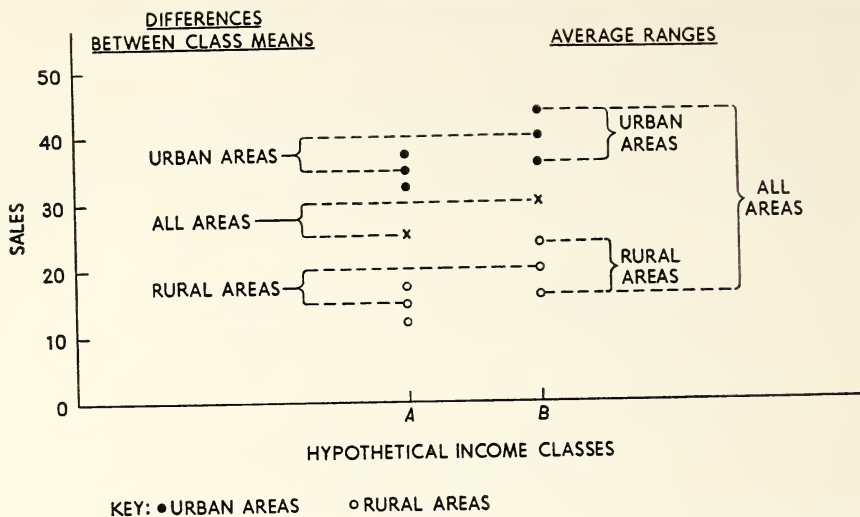


FIGURE 2. Loss of discriminating power due to combining data for urban and rural areas.

estimate of (A-B) is much larger for the pooled data than in the true array. Pooling has reduced the amount of information on the effect of income upon sales that is available to the analyst.

Cross-classifying by all of the relevant variables has the same effect as does matching in experimental designs. The object is to hold the value of one variable constant, so that the relationship between two others can be observed in isolation. The use of matched groups in experimentation eliminates the effects of all the variables that are common to both members of each (matched) pair of observations. The same result is achieved in cross-classification by sorting on relevant variables after the data have been collected.

2. The validity of the balancing process requires that the variable whose impact is being balanced out be structurally independent of the other independent variables. This condition is equivalent to an *absence of interaction*, where the idea of interaction is taken from the context of experimental design. Two independent variables are said to interact if the size of the effect of one depends upon the value of the other. The following simple model for the prediction of Y contains an interaction term:

$$Y = aX + bZ + cXZ$$

The effect upon Y of a change in X depends upon Z because of the term (cXZ).

The validity of the balancing process in cross-classification requires that the value of c be zero; that is, no interaction terms can appear in the model. Figure 2 exhibits the required independence, since the differences

between average sales for the two income classes are the same for both urban and rural areas. That is, sales in both types of areas are effected by income in the same amount, even though their absolute magnitudes are very different.

We emphasize that it is the underlying relationship (which is to be estimated) that must be void of interaction, and not necessarily the sample itself. Sampling or experimental errors could easily cause the change in sales from an actual sample to be slightly different for urban and rural areas, even if the independence assumption is valid for the underlying relationship.

If an important variable must be excluded from the cross-classification because data on it are not available, conclusions based upon the analysis must be somewhat qualified. Moreover, one must proceed with caution at all times because there is always a possibility that some variable has been overlooked. The analyst must ask himself whether the observed effects are reasonable in the light of his prior knowledge of the problem. He must judge whether or not important variables have been excluded. If so, are they likely to be in balance? Is the independence-of-effects assumption reasonable for them? While the data themselves may provide clues, as will be discussed below under the heading of regression, the conclusions reached will stand or fall on the subjective judgment of the analyst and his consultants.

Drawing conclusions from numerical data requires careful work on the part of someone who is familiar with the problem area and who is capable of a high order of subjective reasoning. And yet how many students—and would-be statisticians as well—think only of the mechanical details of computation and presentation? We have seen that even for a relatively simple technique such as cross-classification, the effects of important excluded variables may (1) reduce the amount of information about the desired relationship that is extracted from the data, or (2) actually obscure the true relationship and lead to false conclusions. Only by careful reasoning within the context of the problem can one be reasonably confident that all of the relevant variables have been specified.

CORRELATION ANALYSIS

Two variables are said to be highly correlated if the degree of *linear dependence* between them is large. The correlation coefficient, denoted by the symbol r_{XY} is a summary measure of the association between two variables, X and Y , that is present in a particular set of data. It is a quantitative measure of the extent to which the points in a scatter like that presented in Figure 1 tend to lie on a straight line. If their trend is upward and to the right, as is the case for the sales-income array, $r_{(\text{Sales-Income})}$ will be positive. If the reverse had been true, it would have been negative. The coefficient is defined so that its value must always lie between $+1$ and -1 : it is $+1$ if all the points in the scatter

lie on an upward sloping straight line, and -1 if all of them lie on a line that slopes downward. If X and Y are not linearly related at all, r_{XY} will be equal to zero.

Simple Correlation. The simple correlation coefficient between X and Y is defined:

$$r_{XY} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\left[\sum_{i=1}^n (X_i - \bar{X})^2 \right] \left[\sum_{i=1}^n (Y_i - \bar{Y})^2 \right]}}$$

The capital Greek letter sigma (Σ) is the symbol for addition: the numerator signifies that the quantity $(X_i - \bar{X})(Y_i - \bar{Y})$ for each of the n observations in the scatter ($i = 1, \dots, n$) should be added together. The symbols $\bar{X}$ and $\bar{Y}$ refer to the sample averages of X and Y , respectively.⁷

The formula given above is equivalent to the following expression:

$$r_{XY} = \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \sigma_Y^2}}$$

where σ_X^2 and σ_Y^2 are the variances of X and Y . The quantity in the numerator is a measure of the amount of *covariation* that is present in the sample; it is defined in a manner similar to the variance:

$$\sigma_{XY} = \sum_{i=1}^n \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

Covariance is highly positive if X_i and Y_i values that deviate from their means in the same direction (that is, both positive or both negative) are associated with one another. It is negative if X_i and Y_i for the same observation tend to deviate in opposite directions. In addition, covariance will increase (either in a positive or negative direction) as the variance of X or Y increases.

The correlation coefficient is really a standardized measure of covariation: covariance, σ_{XY} , is divided by the square root of the variances for the particular sample. Standardization allows two correlations to be com-

⁷ For actual computations it is more convenient to use the formula:

$$r_{XY} = \frac{\Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{n}}{\sqrt{\left[\Sigma X^2 - \frac{(\Sigma X)^2}{n} \right] \left[\Sigma Y^2 - \frac{(\Sigma Y)^2}{n} \right]}}$$

where all the necessary quantities may be accumulated on the dials of a desk calculator. Cf., Wilfred J. Dixon and Frank J. Massey, Jr., *Introduction to Statistical Analysis* (New York: McGraw-Hill Book Co., Inc., 1951), pp. 20-21, and 165-69.

pared without regard to the amount of variation exhibited by each variable separately.

A number of scatter diagrams and correlation coefficients are presented in Figure 3. Study of these diagrams will develop some feel for the magnitude of the correlation coefficient, given a visual assessment of the

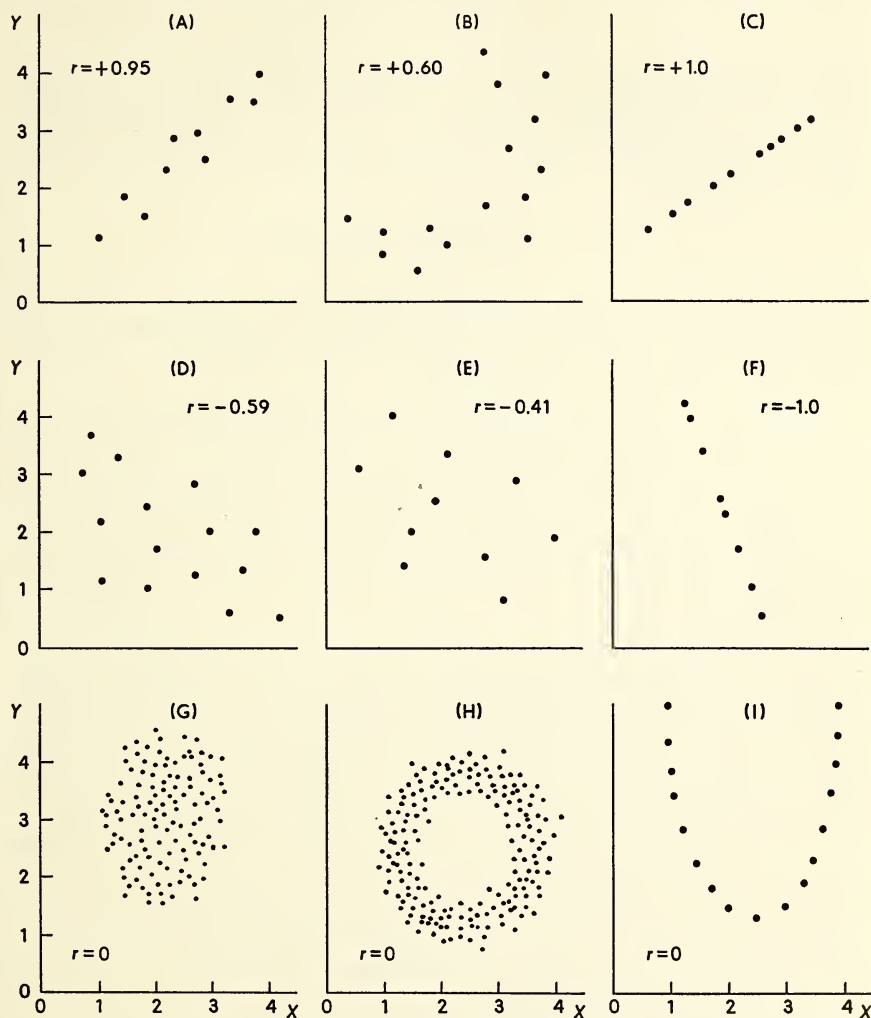


FIGURE 3. Scatter diagrams and correlation coefficients.

degree of association in a particular scatter. Two other facts can be seen as well: (1) the value of r is not affected by the average values of X and Y —only deviations from means enter the correlation formula; and (2) scatters that differ in their direction of trend, but are identical in degree of association (for example, diagrams C and F) yield correlation coefficients that differ only in sign.

Diagrams G through I demonstrate the fact that r_{XY} is a measure of linear association only. While anyone would agree that no relationship between X and Y exists for Diagram G, this can hardly be the case for H and I. The points in H all fall between two concentric circles—hardly a random layout. In I the points lie exactly on the parabola

$$Y = 1.3 + 1.2 (X - \bar{X})^2$$

The relationship is exact, but the correlation coefficient is zero because there is no linear term in the parabola! Correlation coefficients measure only the degree of *linear* association that is present in a set of data. If the equation had been:

$$Y = a + b (X - \bar{X}) + c (X - \bar{X})^2$$

r would have been nonzero because of the linear term, $b(X - \bar{X})$. The correlation would not be perfect; even though the relationship is exact and a linear term is present, the deviations from a linear trend caused by the squared term will have their effect.

The square of r_{XY} is equal to the proportion of the variation of Y and X that is accounted for by the linear relationship between them. In Figure 3-A, the scatter tends along an imaginary line moving upward and to the right. We would say that the line accounts for r^2 , or 90 percent of the variation of X and Y . In diagram C, where all the points lie exactly on the line, 100 percent of their variation is accounted for by the linear relationship.

Correlation coefficients are useful because they summarize the co-variation of variables in large bodies of data. All the information about the amount of linear association between sales and income that is contained in the 18 points of Figure 1, for example, is also contained in the single number r . Such "boiling down" of data into a more concise and easily usable form usually pays dividends in the success of later analysis.

Three Variables. The concept of correlation can be extended to situations where relationships involving more than two variables are to be analyzed. Figure 4 presents a three dimensional scatter diagram. The location of each point ($\cdot$) in the three dimensional space is determined by projecting lines outward from the corresponding points ($\times$) in each of the two dimensional side planes.

A simple correlation coefficient summarizes the amount of linear relationship between *two* variables. Therefore, the *three* dimensional scatter diagram allows the computation of three different simple correlations, since three variables can be divided into pairs in three possible ways. In Figure 4, the simple correlation between Y and X is represented graphically by the scatter of points in the Y - X side plane. Information about the effects of Z is thrown away when the Y - X scatter is considered by itself.

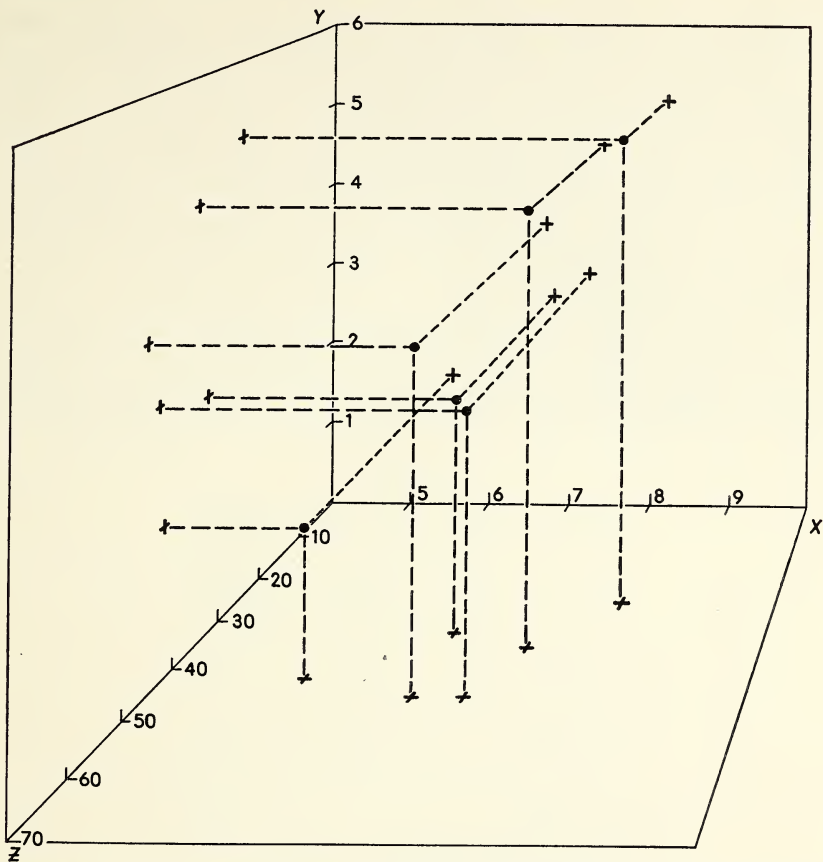


FIGURE 4. Scatter diagram of sales, income, and temperature for Warm brand canine coats (first six observations only). Based on data from Table 1.

The three simple correlation coefficients are usually arrayed in the following form:

	y	x	z
y	1.0	r_{yx}	r_{yz}
x		1.0	r_{xz}
z			1.0

Only the coefficients above the diagonal need be computed, since r_{ij} equals r_{ji} for all possible pairs. The "ones" down the diagonal represent the correlation of each variable with itself. The reader can verify that for a set of N variables (rather than three, as in the present case) there are always exactly $(N^2 - N)/2$ simple correlations to be computed.

The table of coefficients presented above is called the *simple correlation matrix*. As was the case for a single correlation coefficient, the matrix as a whole summarizes all of the information about the degree of linear dependence between the three variables that is contained in the original data. It forms the basic input for regression, discriminant, and factor analysis—the multivariate statistical techniques that will be subsequently discussed. In addition, careful inspection of the simple correlation matrix can sometimes give preliminary information about the way in which the variables are interrelated.

Partial Correlation. The simple correlation coefficient is a summary measure of the degree of linear association between two variables. We now wish to explore the concept of the correlation between two variables when the effects of other variables have been removed. The addition of

TABLE 5
SIMPLE CORRELATION MATRIX FOR MARKET
ANALYSIS OF *Warm Brand Canine Coats**

	Sales Y	Income X	Temperature Z
Sales Y	r_{YY} 1.0	r_{YX} +0.89	r_{YZ} -0.45
Income X		r_{XX} 1.0	r_{XZ} -0.57
Temperature Z			r_{ZZ} 1.0

* Based on the data in Table 1.

the third hypothetical variable given in Table 1, winter temperature, to our analysis of the relation between canine coat sales and income in urban areas will illustrate the point.

Table 5 presents the simple correlation matrix for sales (Y), income (X), and January minimum temperature (Z)—a measure of the winter climate for each metropolitan area. The value of r_{YX} summarizes the linear association between sales and income; it is large and positive, as would be expected from the scatter diagram in Figure 1. The value of r_{YZ} (-0.45) suggests that canine coats are not as eagerly sought in areas having warmer winters, while r_{XZ} (-0.57) shows that the cities in the sample with high income levels tend to have relatively cold winters.

We now ask ourselves whether the combined effect of income and temperature upon sales is really as large as would seem to be implied by the simple correlations r_{YX} and r_{YZ} . If this were the case, the two independent variables should explain $r_{YX}^2 + r_{YZ}^2$ or: $(0.89)^2 + (-0.45)^2 = 0.995$, or 99½ percent of the total variation in sales. But this conclusion would be overly optimistic. Our data for income and temperature are, in

part, measures of the same thing: to be precise, exactly r_{XZ}^2 or 32 percent of the information contained in one is also contained in the other. To claim that 99½ percent of the variation in sales among metropolitan areas can be attributed to the combined effects of income and temperature would require that the effect common to both be counted twice. Since claiming the same result twice would be cheating, some other method for assessing the combined contribution of the two independent variables must be found.

The proper approach is to find a measure of the relationship between income and sales, with the effect of variations in temperature taken out, and vice versa. The measures we seek are called *partial correlation coefficients*. The first is designated by $r_{YX.Z}$ which is read "correlation between Y (sales) and X (income) with Z (temperature) held constant"; and the second by " $r_{YZ.X}$ " where income is held constant. Their values are shown in Table 6 (the method of computation will be discussed below).

TABLE 6
PARTIAL CORRELATION MATRIX FOR MARKET
ANALYSIS OF *Warm Brand Canine Coats**

	Income X	Temperature Z
Sales Y	$r_{YX.Z}$ +0.85	$r_{YZ.X}$ +0.16
Income X	—	$r_{XZ.Y}$ -0.41

* Computed from the simple correlation coefficients in Table 6.

Since partial correlation coefficients measure the relationship between the pair of variables listed ahead of the "dot" as it would appear if the third variable had been held constant, it is safe to regard $r_{YX.Z}^2$ as the *net* contribution of X to an explanation of the variation of Y . Therefore, the combined effects of income and temperature account for exactly $r_{YX.Z}^2 + r_{YZ.X}^2$, or: $(0.85)^2 + (0.16)^2 = 0.748$, or 75 percent of the total variation in sales.

The two partial correlations allow us to disentangle the separate effects of the independent variables. For canine costs, the figures show that, taken by themselves, milder winters are associated with more demand—perhaps because more delicate breeds of dogs reside in the warmer climates. (We re-emphasize that these data are purely hypothetical!) The value of $r_{YZ.X}$ is +0.16, which accounts for only about 2½ percent of the variation in sales, but is positive. The negative simple correlation between Y and Z was really due to the effect of income, which completely swamped that of temperature. On the other hand, r_{YX} is almost equal to

$r_{YX.Z}$: Z is a relatively unimportant cause of variation in Y and so did not affect the other simple correlations very much.

Geometric interpretations of simple and partial correlation coefficients are presented in Figure 5. The points ($\cdot$) in the three dimensional space are projected horizontally to form scatters ($\times$) on each of the three vertical planes. The scatter on plane "0," at the right, includes the projections of *all* the points and so represents the simple correlation between Y and X . Planes 1 and 2 differ from plane 0 because each of them contain the projections of only those points which lie in a specified range of Z , whereas plane 0 embraces all the points without regard to

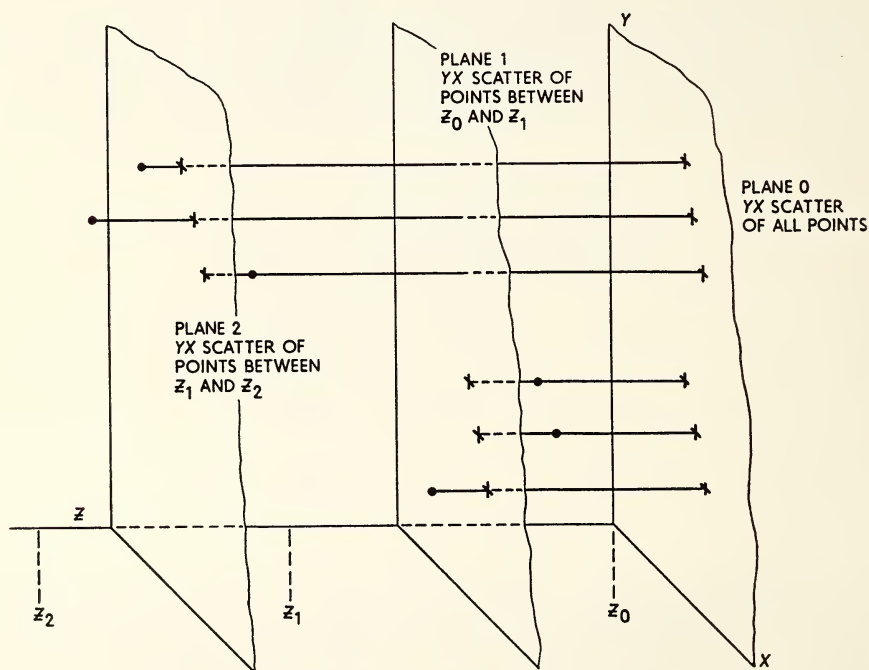


FIGURE 5. Partial correlation scatter diagram (hypothetical data.)

the value of Z . The scatters on planes 1 and 2 also represent simple correlation coefficients; they might be described as follows:

Plane 1:

$$r_{YX.} \text{ (with } Z \text{ held between } Z_0 \text{ and } Z_1 \text{)}$$

Plane 2:

$$r_{YX.} \text{ (with } Z \text{ held between } Z_1 \text{ and } Z_2 \text{)}$$

A crude measure of the relationship between Y and X with the effect of Z removed could be obtained by averaging these two values of $r_{YX.}$ (There is no reason why the two values should be equal.) The reader may note that in this particular example the correlations represented on

both planes 1 and 2 are exactly -1 , whereas the simple correlation on plane 0 is relatively small because the perfect negative relationship between Y and X has been obscured by the effect of Z .

Now let us imagine that more and more planes are added to Figure 5, with the ranges of Z from which points are projected on to each becoming smaller and smaller. As the number of planes becomes very large, the influence of Z upon the value of $\bar{r}_{YX}$ obtained by averaging over the planes becomes negligible, and $\bar{r}_{YX}$ converges to the partial correlation coefficient $r_{YX.Z}$.

The partial correlation of Y and X can be computed directly from the simple correlation matrix by the formula:⁸

$$r_{YX.Z} = \frac{r_{YX} - r_{XZ}r_{YZ}}{\sqrt{(1 - r_{XZ}^2)(1 - r_{YZ}^2)}}$$

The reader should be able to write similar formulas for computing $r_{YZ.X}$ and $r_{XZ.Y}$.

Multiple Correlation. The idea of a multiple correlation between three or more variables has already been discussed indirectly. Designated by the symbol, $R_{Y.XZ}$, it is a measure of the degree of linear relationship between a dependent variable ($\dot{Y}$) and two (or more) explanatory variables (X and Z). As for simple and partial correlations between pairs of variables, its square is equal to the proportion of the variation of Y that is explained by the linear relationship—which in this case involves the two other variables simultaneously.

One method for calculating its value is to compute:

$$R_{Y.XZ} = +\sqrt{r_{XY}^2 + r_{YZ}^2 - r_{XZ}^2}$$

This formula was utilized in the last section, when we wished to calculate the combined contribution of income and temperature to an explanation of variations in sales. We found $R_{Y.XZ}^2$ to be 0.75, which gives a multiple correlation of 0.87.

The multiple correlation coefficient can be viewed in a slightly different way, as well. Let us define a new variable, W , as the *linear combination* of our two independent variables, X and Z . (A linear combination of two or more variables is merely their weighted sum.) For each observation ($i = 1, \dots, n$) in the sample we have:

$$W_i = aX_i + bZ_i$$

We have chosen a *linear* combination rather than some other way of combining X and Z (for example, a quadratic or logarithmic equation)

⁸ George W. Snedecor, *Statistical Methods* (5th ed.; Ames, Iowa: The Iowa State University Press, 1956), p. 430.

because, as the reader will recall, the whole concept of correlation analysis is based upon linear relationships.

Now we will compute the simple correlation coefficient between Y and W in the usual manner. Clearly, the value of r_{YW} will depend upon the particular weighting factors chosen in computing W in the first place: if a is zero, for example, we would have $r_{YW} = r_{YZ}$, and similarly for b . If we were to calculate r_{YW} for a great many different values of a and b , we would find that one pair yields a larger correlation between Y and W than does any other; *this maximum value for r_{YW} is equal exactly to $R_{Y.XZ}$, as computed from the partial correlations of Y with the original independent variables.*

The idea that the effects of a number of independent variables can be summed up in a new variable, W , is of key importance. It will be pursued in the next section, under the heading of multiple regression analysis.

Partial and multiple correlation analysis can be extended to include any number of variables. For partial correlations, the symbol $r_{YX.ZQRL}$ would be read "correlation between Y and X , with the variables Z , Q , R , and L held constant." The formula for its calculation is analogous to that for $r_{YX.Z}$: it explicitly involves all of the possible simple correlation coefficients. Likewise, the multiple correlation coefficient $R_{Y.XZQRL}$ is a measure of the relationship between Y and all of the independent variables taken together. It might be rewritten as r_{YW} , where W is the linear combination of X , Z , Q , R , and L that yields the greatest possible association with Y .

REGRESSION ANALYSIS

Suppose that our hypothetical canine coat firm wanted to estimate the number of sales that could be obtained from a metropolitan area that was not in its original sample. Such an estimate would be required if the firm was considering an expansion of operations into a new market area, for example, Kansas City.

If we had no statistical information whatsoever about Kansas City but believed that conditions there were about like those in our existing markets, our best forecast would be the average level of sales that had been obtained elsewhere. If we designate sales in Kansas City by Y_* , we have

$$\hat{Y}_* = \bar{Y}$$

where the caret ($\wedge$) over Y_* indicates an *estimated* rather than an *observed* value.

This simple estimate of Y_* would be efficient if we really had no information about the new area, but it is inappropriate if Kansas City's income level is known. Since we have already shown that there is a relationship between sales and income in our existing markets, we should

surely take advantage of this knowledge when preparing sales estimates for new areas.

Introduction to Univariate Regression: Forecasting. How should information on income be incorporated into the forecast of sales in Kansas City? The correlation coefficient between the two gives the *magnitude* of linear relationship, but that is all. We need to know the average change in Y that has been associated with a change in the value of X . This figure is the slope of the linear relationship, and is sometimes called the *regression of Y on X* , or more simply the *slope parameter* of the regression equation.

The forecast value for sales in Kansas City, or for any metropolitan area, can be written as the linear combination of the mean of Y (that is, the simple forecast value discussed above) and the difference between the area's X value and the mean of all the X 's:

$$\hat{Y}_* = \bar{Y} + b(X_* - \bar{X})$$

The goal in making a forecast is to make the difference between the forecast estimate and what eventually turns out to be the actual value be as small as possible. This criterion provides the key for obtaining a value for b . Write the forecast error as:

$$u_* = \hat{Y}_* - Y_*$$

If forecasts were prepared for many metropolitan areas, one would hope that the values of u would all be small, clustered around zero, and devoid of any regular pattern. The characteristics of the u 's are so important for the validity of regression procedures that a whole section will be devoted to them.

It will be recalled that the amount of information contained in an unbiased statistical measure is inversely proportional to its variance: therefore the information contributed by $\hat{Y}_*$ will go up as the variance of u_* goes down. Consequently, we must find the value of b that minimizes $\sigma_{u_*}^2$. While u cannot be calculated for Kansas City because actual sales there are not yet known, observed sales figures *are* available for the original group of areas. Various values of b can be used to prepare hypothetical sales estimates for existing market areas and the value of u calculated for each of them. Since the value that yields the smallest sum of squares for u is chosen as the regression slope parameter, the method has become known as *least squares analysis*.

The procedure is shown graphically in Figure 6. The value of b must be chosen in a way such that the sum of the squares of the vertical distances between the raw data points and the regression line is as small as possible. It is apparent, for example, that the sum of squared deviations is larger for the dashed regression line than for the solid one, thus mak-

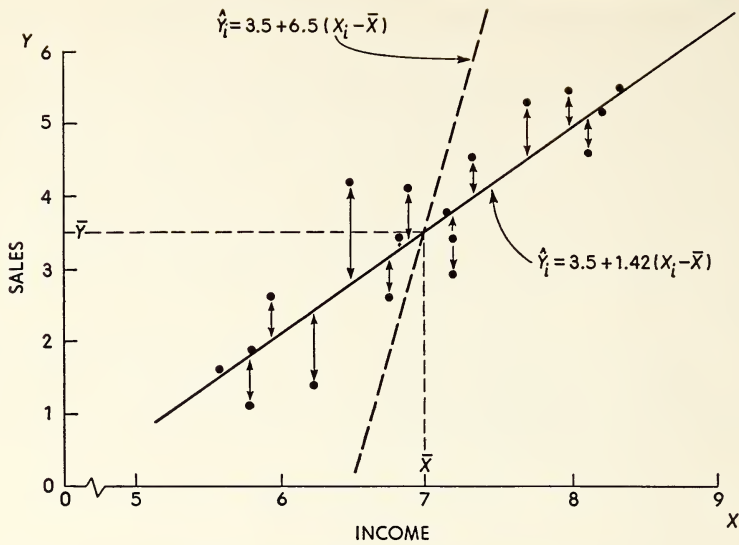


FIGURE 6. Regression of sales (Y) on income (X).

ing the value 1.42 a better estimate of b than is 6.5. Note also that the regression line passes through the point $(\bar{X}, \bar{Y})$; this condition results automatically from the least squares procedure.

It is not necessary to actually compute the squared deviations from a large number of alternative regression lines. The problem has been solved mathematically and the following formula obtained:⁹

$$b = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{\sigma_{XY}}{\sigma_X^2}$$

It can be read as “covariance (XY) over variance (X).” Recalling the formula for the simple correlation coefficient between X and Y , we can write:

$$b = \frac{r_{XY}\sigma_X\sigma_Y}{\sigma_X^2} = r_{XY}\frac{\sigma_Y}{\sigma_X}$$

The regression parameter b represents the value of the correlation coefficient with information on the relative variability of X and Y put back in. Understanding the regression slope parameter allows a clearer interpretation of the correlation coefficient; it is nothing less than the slope of the regression line expressed in standardized units!

⁹ Anderson and Bancroft, *op.cit.*, chap. xiii.

The nature of linear relationships as discussed in connection with correlation coefficients can now be made clear. *Two variables stand in a relationship to one another if knowledge of the value of one will improve the prediction of the other, in the sense of reducing the variance of the forecast error.* For the linear case, this requires that the value of the regression slope parameter be nonzero: if it were zero, the independent variable would drop out of the forecasting equation.

One final point must be considered: *the slope parameter of the regression of Y on X is not the same as that for the regression of X on Y.* The regression equation is not statistically reversible. To estimate the

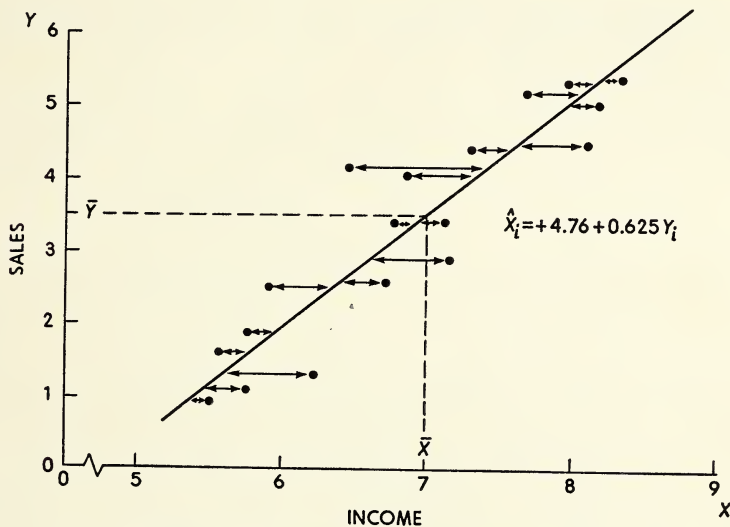


FIGURE 7. Regression of income (X) on sales (Y).

value of income from the number of sales made in a given metropolitan area would require knowledge of b' in the following equation:

$$\hat{X}_i = \bar{X} + b' (Y_i - \bar{Y})$$

A value of b' is optimal if and only if the variance of the forecast error

$$v_i = \hat{X}_i - X_i$$

is as small as possible. Since we are concerned with differences between the X 's, rather than the Y 's as was previously the case, the sum of the squared horizontal deviations between the regression line and the raw data points must be minimized. The model is shown in Figure 7.

The formula for computing b' is similar to that for b . It reduces to:

$$b' = r_{XY} \frac{\sigma_X}{\sigma_Y}$$

In order for the regression relation to be reversible, the condition

$$b' = \frac{1}{b}$$

would have to be satisfied. Substituting the expressions for each coefficient, we would have

$$r_{XY} = \frac{1}{r_{XY}}$$

The equality holds only for the case where $r_{XY} = 1$: all of the points lie on the (single) regression line, and both X and Y could be predicted without any error at all.

Minimizing the sums of squared deviations in the Y direction, as indicated above, yields the *best unbiased forecasts of the dependent variable* that are available from any procedure utilizing linear relationships. This powerful result follows from a fundamental theorem on least squares due to Markov, a Russian mathematician.¹⁰ He has shown that for linear models least squares forecasts are (a) unbiased, and (b) more efficient than those obtained by any other unbiased forecasting procedure.

Univariate Regression: Structural Analysis. Regression was introduced as a technique for the prediction of Y given a particular value of X , but perhaps its major role lies in helping us to understand causal links between variables. In causal analysis, attention focuses on the value of b , which in the example is a measure of the response of sales to changes in income, rather than on the prediction for Y (sales) as in forecasting problems. We might wish to compare the observed sales response to changes in income against our previous beliefs—which may have resulted either from a carefully worked out theory or from a hunch—in order to test our understanding of the environment in which the firm operates. Or, if the effect of income upon sales is itself undergoing change, we may be able to make rough adjustments in the regression slope parameters on the basis of our subjective understanding of the problem area, without waiting for actual data on which to base additional statistical analysis to become available. This can be done only if the effects of income upon sales have been disentangled from those of other variables.¹¹

The regression technique discussed above yields the best possible forecast for Y , but may at the same time fail to give a good estimate of b . Regression can produce good forecasts on the basis of very biased estimates of response coefficients for the independent variables (that is, the b 's).

¹⁰ Discussed in Tintner, *op. cit.*, pp. 83–84.

¹¹ Cf., Hood and Koopmans, *op. cit.*, chap. i.

One of the reasons is already familiar; we considered it under cross-classification and again in connection with partial correlation. Recall the effect of a third variable, temperature, on the net relationship between sales and income. Since income and temperature are negatively correlated, the partial correlation between sales and income as shown in Table 6 is smaller than the simple correlation given in Table 5. Regression coefficients behave in the same way. The value of b computed from the formula given above includes some of the effect of temperature; the true sensitivity of sales to changes in income is somewhat smaller.

To anticipate the section on multiple regression for a moment, we can designate our original value of b as a *simple regression coefficient*, denoted by b_{YX} , and define a *partial regression coefficient*, for which the effects of temperature are excluded, as $b_{YX.Z}$. The latter is analogous to the partial correlation concept, and is read "coefficient of regression Y on X , with Z held constant." For the correlations given in Table 5, we have:

$$b_{YX.Z} = 0.94b_{YX}$$

We would say that the statistic b_{YX} is a *biased* estimator of $b_{YX.Z}$ (although here the bias is small because the effect of temperature on sales is small). A biased estimator is one that, on the average, will give wrong values for the population value of the coefficient under study.

Given the effects of the temperature variable, the value of b_{YX} should not be used as an estimate of the importance of income as a cause of variations in sales among areas. While the error in the example is only in the order of 6 percent, errors of twice or even 10 times as much are not at all uncommon in actual cases. Where more than one variable is believed to exert an effect upon sales, or any other dependent variable, the best course of action is to proceed immediately to the multivariate procedure discussed below.

The *source* of the regression error term, or difference between forecast and actual values of Y , is a second determinant of bias in the estimate of b_{YX} . The error term comes into being through one or both of the following two mechanisms:

1. Not all of the variables that cause fluctuations in the dependent variable have been explicitly included in the regression equation. The combined effects of all the excluded variables can be summed up in the error term, u_i .
2. The measuring process for Y and/or X is subject to random error. The discrepancies between actual and reported values of the regression variables can also be summed up in the error term, u_i .

These are called *errors in equation* and *errors in variables* models, respectively.

The regression procedure described above was derived under the assumption that either (1) or a special case of (2) holds. The error term,

u_i , can arise from the effects of excluded variables, in which case the regression could be written:

$$Y_i \text{ (actual)} = \bar{Y} + b(X_i - \bar{X}) + u_i \text{ (effects of excluded variables).}$$

The u might also be an error of measurement for the *dependent variable*, Y :

$$Y_i \text{ (as measured)} = Y_i \text{ (actual)} - u_i \text{ (measurement error)} = \bar{Y} + b(X_i - \bar{X}).$$

Both models yield the same regression equation, since the quantity u_i can be moved from one side of the equality to the other by simple addition or subtraction.

If the measurement of X is subject to error, on the other hand, the regression equation should be written:

$$Y_i \text{ (actual)} = \bar{Y} + b\{[X_i \text{ (as measured)} + v_i \text{ (measurement error)}] - \bar{X}\},$$

where the error term v_i is multiplied by the parameter b . This result can be recast into the form

$$X_i = \bar{X} - \frac{1}{b}(Y_i - \bar{Y}) + v_i.$$

It is the same as an "ordinary" regression with X rather than Y taken as the dependent variable. The parameter $\frac{1}{b}$ is equal to b' , obtained by minimizing the sum of squares of the horizontal deviations in Figure 7. On the other hand, regression of Y on X where the X 's are subject to error will yield an estimate of b that is too small in absolute value; we say that such estimates of b are biased toward zero. Unbiased parameter estimates can be obtained only if the sum of squares of the error component of the regression equation is minimized: if the measurement of Y contains error we must minimize the Y direction, or if X contains error the squares of the X deviations must be minimized.

If both Y and X are subject to measurement errors, the sums of *diagonal* deviations from the regression line must be minimized if unbiased estimates of b_{YX} are to be obtained. The direction of the proper diagonal is defined by the ratio of the variances of the measurement errors for X and Y .¹²

Errors due to excluded variables and measurement errors for the dependent variable can occur at the same time without invalidating the normal regression procedure. The analyst can also take comfort from the

¹² Laurence R. Klein, *A Textbook of Econometrics* (Evanston, Ill.: Row, Peterson & Co., 1953), pp. 282-305.

fact that errors of measurement for X that have small variance relative to σ_u^2 will lead to only a small bias in the estimate of b . In addition, the reader will recall that forecasts of the dependent variable are not affected by the source of the regression error term.

Review of Bias and Efficiency in Regression. Questions of *bias* refer to the average or *expected value* of an estimate: will $\hat{Y}$ be right on the average, in the sense that overestimates will be canceled out by underestimates? Questions of *efficiency*, on the other hand, are based upon the inherent variability of an estimate.¹³ One forecast is more efficient than another if it has a smaller *variance*, and it is quite possible for one unbiased estimate to be more efficient than another. Since we are looking for forecasts that have errors which are small and distributed irregularly, it is clear that estimates that are both unbiased *and* efficient are desired.

The coefficient b_{YX} is a biased estimator for $b_{YX,Z}$, the true response of sales to income. Nevertheless, the regression equation:

$$\hat{Y}_i = \bar{Y} + b_{YX}(X_i - \bar{X})$$

will yield an unbiased estimate of Y_i regardless of whether or not the additional variable Z should have been included. Bias in our estimate of the effect of income will not affect the forecast of Y as long as the relationship between X and Z remains unchanged. This can be demonstrated in the following way: Z is not included in the regression, but X can be used as an estimator of Z , as well as a measure of itself; therefore, the estimated value of Y is based on information on X , which is available directly, and upon Z , which is introduced via X and the underlying relationship between X and Z . Only a part of the effect of Z upon Y can be determined as long as the correlation between X and Z is not perfect. Consequently, the efficiency of the forecast of Y can be improved by adding Z to the regression equation.

Assumptions Underlying the Use of Regression. We are now in a position to systematically state the assumptions that underlie the use of regression methods for forecasting and structural analysis.¹⁴ Five assumptions will be given. Assumption (4) is a key factor in structural analysis, but does not apply to forecasting. All the others apply to both.

1. The most basic assumption for all kinds of regression is that the regression error term must be *randomly distributed*. The reasons for this

¹³ For an elementary discussion of bias and efficiency in the context of sampling theory, see Spurr, Kellogg, and Smith, *op. cit.*, pp. 232-33. A more advanced treatment can be found in Anderson and Bancroft, *op. cit.*, chap. viii.

¹⁴ An excellent and highly readable summary of assumptions underlying regression and similar statistical methods can be found in Stefan Valavanis, *Econometrics—An Introduction to Maximum Likelihood Methods* (New York: McGraw-Hill Book Co., Inc., 1959), pp. 8-17.

requirement are familiar already: they are the same as those discussed under the heading of randomization in experimental design.

The requirement is fulfilled where data from a properly designed experiment are being analyzed. For observational data, measurement errors due to sampling for the dependent variable are usually distributed randomly. (The independent variables must be measured exactly if the assumptions underlying ordinary regression analysis are to be valid.) In errors in equations models, on the other hand, the assumption will normally be justified only if all the variables whose *individual* effect on Y is large are included as independent variables in the regression.

If we were to discover that one of the metropolitan areas in our canine coats sample charged a \$100 annual dog licensing fee, for example, the validity of the regressions calculated above would be open to serious question. Since such a high fee would limit dog registration to upper income families, we could easily predict that sales *per licensed dog* would be much higher for that area than for others with similar income characteristics: the value of u_i would be large because the regression of sales per dog on income would underestimate the true sales figure. Given this situation, we should not regard the error as being randomly distributed.¹⁵

We also proceed as if the errors have *zero expected value* (for example, the measurement of Y was unbiased), although the failure of this assumption will affect only the estimate of the mean of Y , and not that of the regression slope parameter.

2. The second assumption deals with the variance of the error term. We must postulate that the range of variation between actual and forecast values of Y is no more likely to be large for one metropolitan area than for another. If we think that the people of Boston have the same average characteristics as do those in Philadelphia, but are much more variable in their day to day behavior, we should not assume that forecast errors for the two cities will have the same variance. In technical terms, we say that the u 's must be *homoscedastic*.

Heteroscedasticity does not imply that our structural estimates or forecasts will be biased; on the contrary, uniform error variances are required only if highly efficient estimates are desired. The reason can be seen easily for models where the u 's occur because of errors in the measurement of Y , although the result holds for error in equation models as well. Suppose (a) the number of sales for Philadelphia was known exactly (without error), (b) sales for Boston were known only approximately, and (c) sales for all the other cities were known with an accuracy between that of Philadelphia and Boston. It would make sense to give

¹⁵ In the context of experimental design, this u_i could be considered as randomly distributed if the metropolitan areas in the sample had been randomly selected in the first place. Random selection is unlikely for observational data of the kind considered in this example, however.

Philadelphia, for which accurate information is available, more weight in estimating the regression equation. Likewise, Boston should have a smaller influence on the estimates than cities whose Y measurements have smaller variances. By using the ordinary least squares method we ignore these differences, and hence throw away information about the relative accuracy of the observed Y 's; this loss of information results in forecasts and estimates of the regression slope parameter that are less than fully efficient.

3. The u 's must be statistically independent of one another, in the sense that knowing the value for any u_i will add no information about the value of u_j (where $i \neq j$). Measurement errors for different Y 's in errors in variables models, and excluded variables in errors in equations models, must not be correlated with one another.

This assumption is closely related to (1), since randomly distributed errors are the most likely to be statistically independent. We must recognize, however, that two truly random variables may be highly correlated: for example, the total number of "heads" obtained in a coin tossing experiment after trial t will be highly correlated with that after trial $t + 1$, although both are random variables. Regressions on time series data often have error terms that exhibit exactly the same characteristics as demonstrated in the coin tossing experiment; we would say that these errors are *autocorrelated*.¹⁶ For survey data, the assumption of independent errors may be seriously questioned in cases where the observations are obtained by sampling clusters of respondents: for example, errors in equations which attempt to explain the behavior of next-door neighbors are quite likely to be correlated.

Failure of the independence assumption does not affect the expected values of forecasts or regression parameters, so that we can be assured that no bias will be introduced by autocorrelation of the error terms. Efficiency does suffer, and what is potentially more serious is the fact that estimates of the *forecast variance* and the *variance of the regression slope parameter* (denoted σ_b^2) will in general be biased downward. If the regression errors are highly correlated among themselves, the computed variances will indicate more information on b than we actually have. This misleading statement is often more dangerous than the loss of efficiency itself, and especially so since no adequate means of correction exist.

4. Structural analysis (but not forecasting) requires that the u 's be uncorrelated with the independent variable, X , in the regression equation. While this assumption usually holds for errors in variables models, it must always be challenged and considered carefully when dealing with errors in equations.

¹⁶ Cf., Gerhard Tintner, *Econometrics* (New York: John Wiley & Sons, Inc., 1952), chap. x.

By requiring that r_{uX} be equal to zero, we are doing nothing more than formalizing the conditions we discussed under cross-classification, and again with respect to structural estimation and bias in regression. For the former, cell by cell balancing is a device for setting the sample correlation between X and a particularly important excluded variable to zero. For regression, we found that $b_{YX} \neq b_{YX.Z}$ for $r_{XZ} \neq 0$. Since Z was lumped into the error term in the single variable regression equation, a nonzero r_{XZ} violated assumption (4) and produced a biased estimate of the slope parameter.

Failure of this assumption can lead to serious bias in the estimate of b , and the analyst must be constantly on the lookout for variables that may: (a) affect the dependent variable, (b) be correlated with an independent variable, and (c) are not included in the regression equation. But one never knows when another significant variable may be lurking unseen in the shadows, and the bias introduced will never show up in the variance of the parameter estimate ($\hat{\sigma}_b^2$).

While the statistician has no assurance that his results are not biased and misleading, the *proximity theorem* of regression offers considerable comfort. It states:¹⁷ (a) if the correlation between u and X is small, the bias of the slope parameter estimate will be small; (b) if the variance of u is small, the bias will be small; and (c) if both r_{uX} and σ_u^2 are small, the bias will be negligible. If one makes an honest and informed attempt to include all of the relevant variables in the regression equation, the parameter estimates are likely to be relatively unbiased.

5. Our final assumption refers to the *linearity of the model* upon which regression forecasts or structural estimates are based. Do we believe that the true relationship between Y and X is strictly linear (or can be made so by a suitable transformation of variables),¹⁸ or are we approximating some other form of relationship with a linear regression equation? Such approximations are probably more the rule than the exception in regression analysis, but it is important to recognize that departure from a strictly linear model requires a restriction on the distribution of the *independent* variable.

The theory of regression discussed in most textbooks is based upon strictly linear models.¹⁹ Given a linear relationship, the values of the independent variables can be selected arbitrarily, and the value of Y measured for each. While often applied to observational data, this approach is

¹⁷ Herman Wold, *Demand Analysis* (New York: John Wiley & Sons, Inc., 1953), pp. 37-38.

¹⁸ The concept of regression linearity refers to the way in which the coefficients enter the equation. For example, $\log Y = a + b \log X$ is considered to be a linear regression, while the algebraically equivalent $Y = A \cdot X^b$ is not. Likewise, $Y = a + bX + cX^2$ is linear as far as the mechanics of regression are concerned, because the variables (X) and (X^2) enter in linear combination.

¹⁹ Wold provides a notable exception. *Ibid.*, chap. xii.

particularly well suited to the analysis of experimental results, where the behavior of the independent variables has been specified by the experimenter.

Arbitrary selection of the X values can lead to difficulties in the comparison of different regression lines where the linear regression equation is only an approximation to a more complicated relationship. Figure 8 shows how different linear approximations to the same relation will be obtained, depending upon the particular set of X 's appearing in the sample. Parameters of regression "A" were estimated using X 's that were

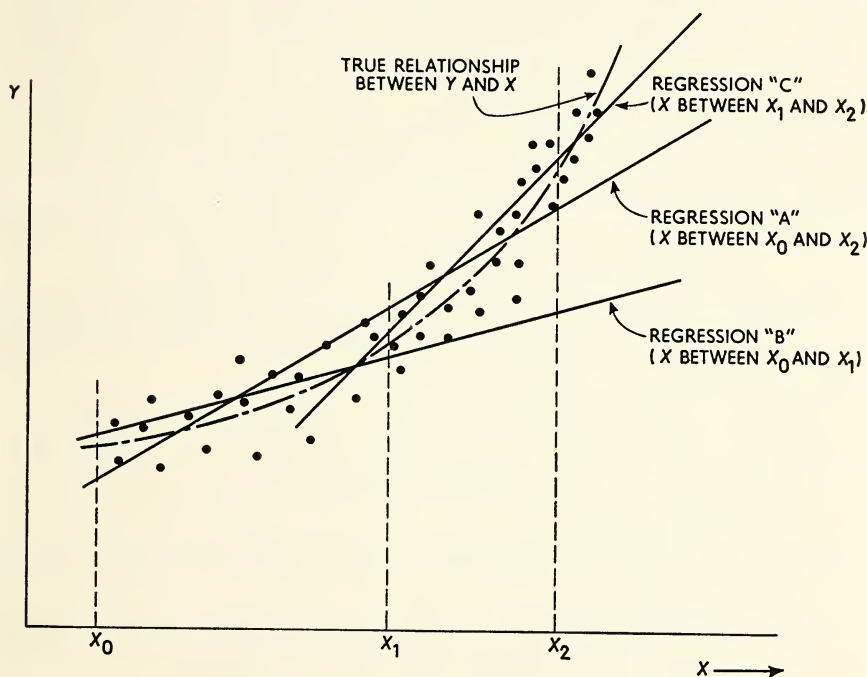


FIGURE 8. Linear regression equations as approximations to an underlying curvilinear relationship.

uniformly distributed within the interval X_0 to X_2 , while lines "B" and "C" are based on X 's in the subintervals X_0 to X_1 , and X_1 to X_2 respectively. While all of the lines are good approximations to the true curve over their respective X -intervals, it would be a mistake to consider all of them as merely regressions of Y on X .

In other words, we must assume that the relationship between Y and X is strictly linear if different sets of arbitrary values of X are to yield comparable regression parameters. If the regression is regarded as being only a linear approximation, the distribution of the independent variable, X , in the sample must be carefully controlled in order for the regression parameters to be interpreted properly.

Consider the following example. Our company has prepared regressions of sales on income, by metropolitan area, during two successive years, and we wish to determine whether the relationship has changed during the period. If we find that the regression slope parameter is much larger for the second year than for the first, but the second sample contained many more areas with high incomes, what should we conclude? If we are very sure that the relationship between sales and income is strictly linear we would disregard the differences between the X distributions for the two years and conclude, on the basis of the two b 's alone, that sales sensitivity to income had increased during the year. If we cannot be sure of the underlying relationship, however, we had better allow for the possibility that the difference between the b 's occurred because of a situation like that shown in Figure 8, that is, that the regressions for the two years approximate different portions of a stable but curvilinear relation. In this case, two samples having the same X 's would allow us to determine whether a change had occurred. A transformation of variables, or the addition of higher order terms in X to the regression (for example, X^2 or X^3) might restore our powers of comparison, but only if a wide range of income variation were available in both samples.

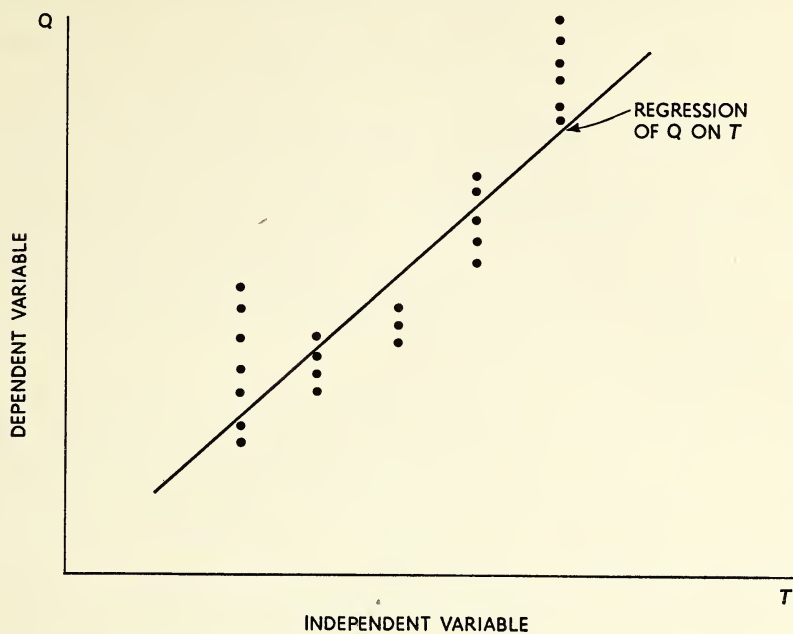
Summary of Regression Assumptions. The five assumptions underlying the use of regression methods for making forecasts and estimating structural parameters are re-stated below:

1. The error must be randomly distributed, with zero expected value.
2. The error variance must be the same for all values of X . (Homoscedasticity)
3. The individual errors, u_i , must be statistically independent of one another. ($r_{u_i u_j} = 0$, for $i \neq j$)
4. The errors must be uncorrelated with the independent variable in the regression equation. ($r_{ux} = 0$) Applies to problems of estimating structural parameters only.
5. The underlying relationship between Y and X must be strictly linear if the regression slope parameter and forecasts based upon it are to be independent of the distribution of the X 's in the sample.

Analysis of the Regression Residuals. Let us focus on the differences between the actual and forecast values of Y , as defined by the vertical distances between the points in the scatter diagram of Figure 6 and the regression line. They are written as r_i , and result from the application of the regression method to a given set of data points; *they are not necessarily the same as the theoretical u_i which are specified as part of the underlying model.* Frequency distributions of the r_i and u_i may differ for one of two reasons: (1) the assumptions upon which the regression method is based were violated; or (2) the regression assumptions were valid, but not enough observations on the r_i were available to produce a good estimate of the u_i distribution. In the latter case, the observed values of r_i can be regarded as a sample from the underlying u_i population.

Figure 9 presents a scatter diagram, regression line, and a graph of re-

A. THE RAW SCATTER DIAGRAM



B. REGRESSION RESIDUALS (r_i)

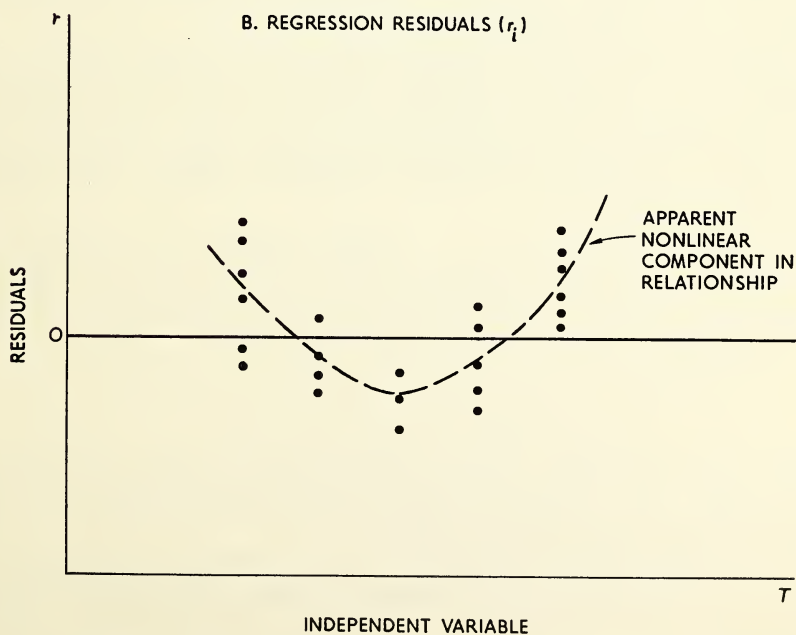


FIGURE 9. Effects of a curvilinear relationship on regression residuals.

gression residuals for two new hypothetical variables, Q and T . Inspection of the residuals plot suggests that at least two of the assumptions about the underlying error process were not justified: (1) the residuals exhibit a definite nonlinear pattern (violation of assumption 5) and (2) the variances of the residuals are not the same for the various values of X (violation of assumption 2). The analyst might therefore consider adding a nonlinear term in X (perhaps X^2) and working with the multiple regression model discussed below; and/ or using a *weighted regression*²⁰ method to compensate for the differing variances in order to improve estimating efficiency. Both moves should be considered carefully, since aberrations in the r_i are often caused only by fluctuations of sampling, and neither of the more complicated techniques can be used without incurring additional costs.

The independence of errors assumption (3) can be roughly evaluated in a similar manner. For time series, one can look at the residuals for adjacent periods, while if the data were obtained from a sample survey, residuals for neighbors, adjacent blocks, cities, and so on can be examined. If a marked degree of clustering of r_i values can be discerned, the validity of assumption (3) is in doubt.

Assumption (4) can never be evaluated using information on the observed residuals, since the regression method insures that the r_i will be uncorrelated with the independent variable whether u_i is or not. While higher order relationships are possible, as is demonstrated by Figure 9, one will always get a correlation coefficient of precisely zero. (Recall Figure 3-I for a demonstration that $r_{rT} = 0$.)

Analysis of the regression residuals often provides a great deal of information for the analyst. Without it there is no way of checking his intuitive assumptions about the nature of the error term. While the available tests are inexact and incomplete, their use is a great deal better than no testing at all. Residuals should be computed, plotted, and analyzed wherever possible.

Multiple Regression Analysis. Multiple regression is an extension of the univariate regression principles to allow the effects of more than one independent variable to be taken into account at the same time. The multiple regression of sales on income and temperature is written:

$$Y_i = \bar{Y} + b_{YX.Z}(X_i - \bar{X}) + b_{YZ.X}(Z_i - \bar{Z}) + u_i$$

This equation has exactly the same interpretation as those in the preceding sections. All of the same conditions and assumptions apply for both structural estimation and forecasting.

The partial regression parameters ($b_{YX.Z}$ and $b_{YZ.X}$) are closely related to the equivalent partial correlation coefficients discussed above.

²⁰ Klein, *op. cit.*, pp. 305-13.

Their values can be computed from the simple correlation matrix and the variances of Y , X , and Z by using the following formulas:

$$b_{YX.Z} = \frac{\sigma_Y}{\sigma_X} \left[\frac{r_{YX} - r_{ZX}r_{YZ}}{1 - r_{ZX}^2} \right]$$

$$b_{YZ.X} = \frac{\sigma_Y}{\sigma_Z} \left[\frac{r_{YZ} - r_{ZY}r_{YX}}{1 - r_{ZX}^2} \right]$$

Note that the simple correlations between all of the pairs of variables appear in both equations.

Two topics need particular emphasis. We will discuss the problem of multicollinearity first, and follow it with an assessment of the practice of substituting two or more univariate regression equations for a single multivariate one. All of the examples will refer to multiple regressions having two independent variables, but the discussion can be generalized directly to cases having a much larger dimension.

Multicollinearity is said to be present in a multiple regression computation if the independent variables are highly correlated among themselves. This condition reduces the efficiency of the estimates for the regression parameters because for given values of r_{YX} and r_{YZ} the amount of information about the effect of each independent variable, taken separately, declines as r_{XZ} increases.

The reduction in efficiency can easily be seen in the limiting case, as r_{XZ} approaches one. If r_{XZ} equals one, we know, for the sample at least, that both independent variables varied in perfect proportion to one another. But if X and Z vary together for all observations in the sample, how can we hope to separate their influence on Y ? X and Z can have distinctly different effects in the underlying population, but our particular sample does not contain enough information to separate them. In terms of the regression coefficients, we see that the formulas given above break down when $r_{XZ} = 1$; no calculation is possible when the denominator ($1 - r_{XZ}^2$) is equal to zero.

It can be shown that the variance of the estimates for both $b_{YX.Z}$ and $b_{YZ.X}$ is directly proportional to the quantity $(1 - r_{XZ}^2)$: their variance is infinite (estimating efficiency is zero) when $r_{XZ} = 1$, and decreases steadily as r_{XZ} declines. Thus, correlation between X and Z reduces the efficiency of estimates of the regression slope parameters. Consequently, it is desirable to design experiments or to use observations whose values of X and Z exhibit as low a correlation among themselves as is possible. Observational data often exhibits high multicollinearity, but sometimes a careful selection of variables or observations can minimize the damage. At other times it may be possible to combine two or more independent variables into a new variable and use that for computing the regression line. Examples of this approach are found in the readings.

The efficiency of *forecasts of Y*, on the other hand, is unaffected by the correlation between the independent variables (unless $r_{XZ} = 1$, where the multiple regression technique breaks down and the univariate method must be adopted). Forecasts depend only upon the total amount of information about *Y* contributed by *X* and *Z*, not on their effects taken separately. While it is desirable to use independent variables that together contribute a maximum of information, the correlation between them does not matter, per se.

Short-cut methods for computing multiple regression coefficients are

TABLE 7
SIMPLE AND MULTIPLE REGRESSION OF *Warm Brand Canine Coats* ON
INCOME AND TEMPERATURE*

A. Calculation of the multiple regression coefficients:

$$b_{YX.Z} = \frac{\sigma_Y}{\sigma_X} \left[\frac{r_{YX} - r_{ZX}r_{YZ}}{1 - r_{ZX}^2} \right] = \frac{1.39}{0.87} \left[\frac{0.89 - (-0.57)(-0.45)}{1 - (-0.57)^2} \right] = +1.51$$

$$b_{YZ.X} = \frac{\sigma_Y}{\sigma_Z} \left[\frac{r_{YZ} - r_{ZX}r_{YX}}{1 - r_{ZX}^2} \right] = \frac{1.39}{14.7} \left[\frac{-0.45 - (-0.57)(+0.89)}{1 - (-0.57)^2} \right] = +0.0086$$

B. Calculation of the simple regression coefficients:

$$b_{YX} = \frac{\sigma_Y}{\sigma_X} r_{YX} = \frac{1.39}{0.87} (+0.89) = +1.42$$

$$b_{YZ} = \frac{\sigma_Y}{\sigma_Z} r_{YZ} = \frac{1.39}{14.7} (-0.45) = -0.043$$

C. Forecast of sales for Kansas City:

$$(\text{Income } (X) = 7.05; \text{ temperature } (Z) = 26.7)$$

Using multiple regression coefficients:

$$\hat{Y} = 3.52 + 1.51(7.05 - 6.96) + 0.0086(26.7 - 40.9) = 5.06$$

Using simple regression coefficients:

$$\hat{Y} = 3.52 + 1.42(7.05 - 6.96) - 0.043(26.7 - 40.9) = 5.68$$

* Computed from the simple correlation coefficients in Table 6.

sometimes utilized to avoid the use of complicated mathematical formulas. While short cuts are useful if employed correctly, they can easily lead to seriously biased results.

The procedure usually works like this: (1) compute the univariate regression of *Y* on *X*; (2) compute the univariate regression of *Y* on *Z*; (3) write the multiple regression equation as:

$$\hat{Y}_i = \bar{Y} + b_{YX}(X_i - \bar{X}) + b_{YZ}(Z_i - \bar{Z})$$

where the estimates of the slope parameters are obtained from the two simple regressions. This method causes substantial bias in the estimates for both *Y* and the effects of *X* and *Z*, as can be seen from the calculations presented in Table 7.

The multiple regression procedure (Part A) yields unbiased estimates of the separate effects of income and temperature. Both are positive, indicating that higher incomes *and* milder winters are associated with more sales of canine coats. These results are consistent with the partial correlations given in Table 6.

The simple regressions (Part B) reproduce the simple correlations between Y and X , and Y and Z ! The estimate of each effect has been biased downward; that for Z even has the wrong sign. The short-cut forecast for Y is strongly biased as well, since information contributed by both X and Z has been counted twice.

Biases of this kind will occur whenever simple regression coefficients are substituted for partial ones in a multiple regression equation, except when the correlation between the two independent variables is equal to zero. Since one almost never finds $r_{XZ} = 0$ in practice, the best rule is:

Always compute the partial regression coefficients when a multiple regression approach is indicated by the nature of the problem.

The advent of electronic computers has made multiple regression computations almost routine. The author has performed multiple regression calculations on 12 variables in less than half an hour on an IBM 1620; on the vastly more powerful IBM 7090, the computations take considerably less than a minute.

There is no longer any excuse for reporting results based on the short-cut method described above, unless it can be shown conclusively that the correlation between the independent variables is negligible.²¹

MULTIPLE DISCRIMINANT ANALYSIS

Multiple discriminant analysis is a statistical technique for making forecasts or estimating structural parameters in problems where the dependent variable appears in dichotomous form (that is, "yes" or "no," "bought" or "did not buy," etc.). Its use and interpretation are much the same as in multiple regression analysis: a linear combination of numerical values for two or more independent variables is used to predict the behavior of a dependent variable.²²

Let us imagine that the forecasting assignment discussed in the previous section is complicated by the lack of quantitative data on the present sale of canine coats in each market area. We assume that accurate numerical information on income and temperature is still available, but that the only data we have on sales are the opinions of our sales manager. This

²¹ A variation on the above theme involves regressing the residuals of the first (simple) regression on a second independent variable. This procedure gives biased estimates for both b 's (where $r_{XZ} \neq 0$) but does not affect the forecast of Y . The reasons are similar to those discussed under partial correlation above.

²² The theory of multiple discriminant analysis is given in Paul G. Hoel, *Introduction to Mathematical Statistics* (New York: John Wiley & Sons, Inc., 1947), pp. 121-26; and Tintner, *op. cit.*, pp. 96-102.

executive has had long experience with the company and is very sure which market areas are "good" and which are "bad" sales producers. The problem is to predict whether a potential new market will eventually fall into the good or the bad category by his judgment.

Regression forecasts of sales are ruled out because numerical data on the dependent variable are not available; but the technique of multiple discriminant analysis can be used to advantage. Our first task is to use existing data on income (X), temperature (Z), and the sales manager's opinions to estimate the parameters of a *discriminant function*:

$$f = c_X X + c_Z Z$$

Once the values of c_X and c_Z have been determined, f can be evaluated for any new market and its value used to predict success or failure. In the example, large values of f will lead to a prediction of good results, and small values will cause us to predict poor sales performance (by the sales manager's standards).

Logic and Computation. The logic of multiple discriminant analysis can most easily be shown with the aid of Figure 10. All of the arguments can be extended to situations where three, four, or more independent variables are to be included in the analysis. The form of the scatter diagram for X and Z is already familiar, but one piece of information has been added: those areas classified as good sales producers are denoted by stars (*), while those classified as bad are shown with circles (○). A *multiple discriminant boundary line* divides the domain of the stars from that of the circles. We wish to find the boundary that discriminates most effectively; of course perfect discrimination by means of a linear boundary is impossible except in special cases. (Why?)

Values of f , as computed from the discriminant function given above, can be represented by the perpendicular distance from each of the points to the discriminant boundary line. Some of the distances are designated by heavy dashed lines in Figure 10. The reader can easily verify that the greater the distance between a point and the discriminant boundary line the greater is the chance that the area will be correctly classified.

The slopes of the dashed lines, and hence of the discriminant boundary line, are determined by the parameters c_X and c_Z in the discriminant function. The parameters can be computed from the sample data by using the following formulas:

$$c_X = \frac{S_{ZZ}d_X - S_{XZ}d_Z}{S_{ZZ}S_{XX} - S_{XZ}^2}$$

$$c_Z = \frac{S_{XX}d_Z - S_{XZ}d_X}{S_{ZZ}S_{XX} - S_{XZ}^2}$$

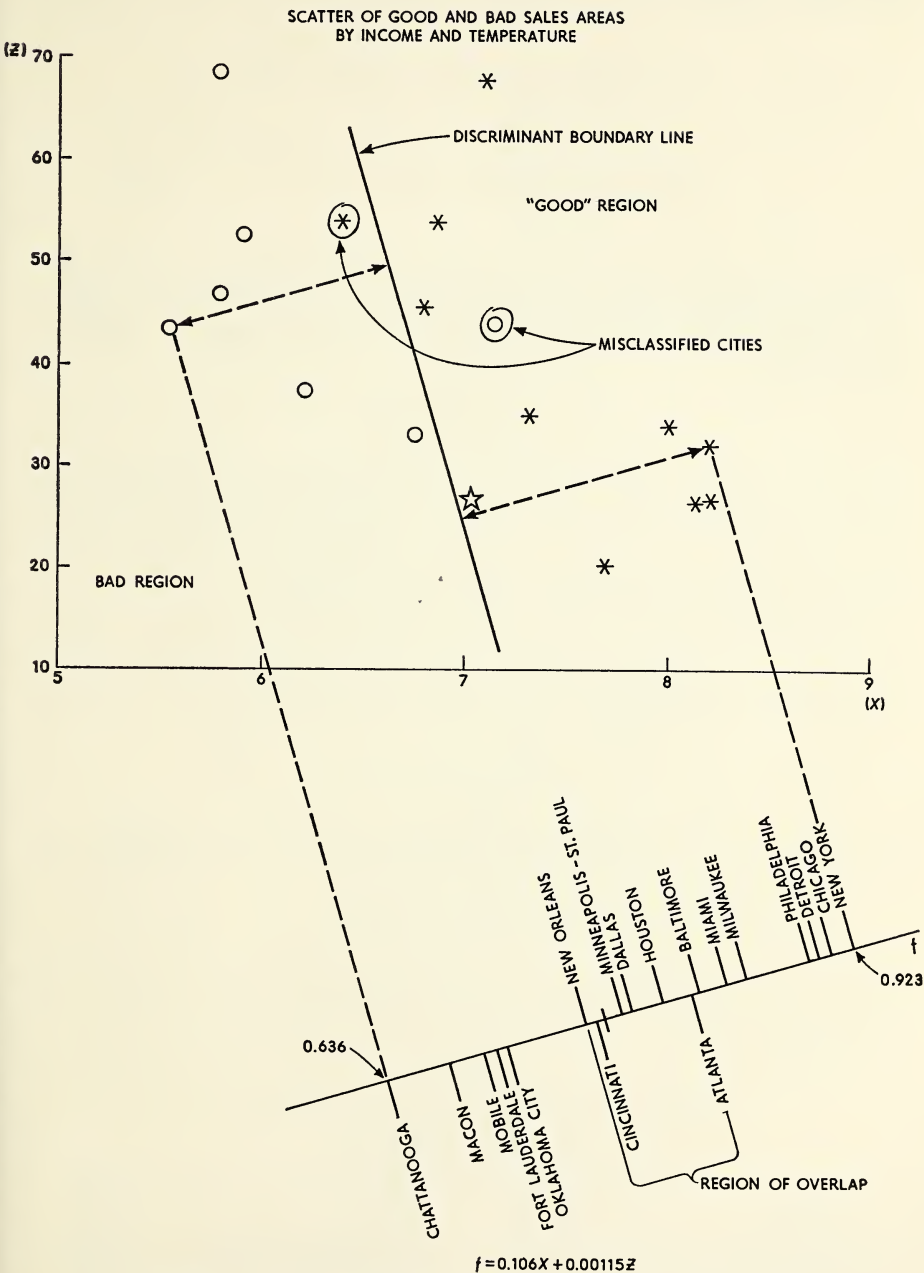


FIGURE 10. Multiple discriminant analysis of Warm brand canine coats.

S_{XX} , S_{ZZ} , and S_{XZ} are the sample moments (they are closely related to the sample variances and covariance for X and Z):

$$S_{XX} = \sum_{i=1}^n (X_i - \bar{X})^2; S_{ZZ} = \sum_{i=1}^n (Z_i - \bar{Z})^2$$

$$S_{XZ} = \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

All sums are taken over the whole sample without regard to whether the observation is on a "good" or "bad" market area. The d 's are the differences between the averages of X and Z for the two types of observations, that is,

$$d_X = \bar{X}^* - \bar{X}^0; \text{ and } d_Z = \bar{Z}^* - \bar{Z}^0$$

Note the similarity between these formulas and those for the partial regression coefficients: the major differences are that the sample moments are used instead of correlation coefficients, and that the values of d_X and d_Z are substituted for the nonexistent r_{XY} and r_{ZY} .

Computations for the canine coats data produced:

$$d_X = \frac{81.97}{11} - \frac{43.13}{7} = 7.45 - 6.16 = 1.29$$

$$d_Z = \frac{408.4}{11} - \frac{324.3}{7} = 37.13 - 46.33 = -9.20$$

$$S_{XX} = 13.6; S_{ZZ} = 3880; S_{XZ} = -129$$

These data yield the multiple discriminant function used in Figure 10:

$$f_i = +0.106X_i + 0.00115Z_i$$

The value of f_i for each of the sample metropolitan areas is given in Table 8, and plotted along the upward sloping lower axis in Figure 10.

TABLE 8
VALUES OF THE MULTIPLE DISCRIMINANT FUNCTION FOR
MARKET AREAS OF *Warm Brand Canine Coats*
($f_i = 0.106X_i + 0.00115Z_i$)

Atlanta.....	0.809	Macon.....	0.664
Baltimore*.....	0.816	Miami*.....	0.834
Chattanooga.....	0.636	Milwaukee*.....	0.839
Chicago*.....	0.896	Minneapolis-St. Paul*.....	0.771
Cincinnati.....	0.752	Mobile.....	0.688
Dallas*.....	0.772	New Orleans*.....	0.746
Detroit*.....	0.888	New York*.....	0.923
Fort Lauderdale.....	0.692	Oklahoma City.....	0.704
Houston*.....	0.789	Philadelphia*.....	0.884

* "Good" sales producing areas.

Cities with good sales records are listed above the f axis, while those with poor marks are below. It is apparent that the f values for the following "good" and "bad" areas overlap:

<i>Good</i>	<i>f</i>	<i>Bad</i>
New Orleans.....	0.746	
	0.752.....	Cincinnati
Minneapolis-St. Paul.....	0.771	
Dallas.....	0.772	
Houston.....	0.789	
	0.809.....	Atlanta

While all areas having an f greater than 0.809 clearly fall into the "good" category, and all those with f less than 0.746 are bad, the areas listed have not been classified effectively. Consequently, we will be uncertain about the proper classification for a new area whose f falls within (or even near) the range of overlap. We may wish to face up to this dilemma directly, and report a "don't know" for the uncertain areas. On the other hand it may be necessary to make a best guess, in which case the following procedure for finding a critical value for the discriminant function appears to be reasonable.

Begin by identifying all of the sample observations that fall within the range of overlap. (This has been done, and the results given above.) Then find the range of critical values of f that results in the *minimum number of misclassifications*. In the example, f 's between (but not including) the indicated values produce the following misclassifications:

MISCLASSIFICATIONS		
<i>Range of f</i>	<i>Number</i>	<i>Areas</i>
0.746-0.752	3	New Orleans, Cincinnati, Atlanta
0.752-0.771	2	New Orleans, Atlanta
0.771-0.772	3	New Orleans, Minneapolis-St. Paul, Atlanta
0.772-0.789	4	New Orleans, Minneapolis-St. Paul, Dallas, Atlanta
0.789-0.809	5	New Orleans, Minneapolis-St. Paul, Dallas, Houston, Atlanta

The best critical value of f must lie between 0.752 and 0.771, where only two areas are misclassified. For simplicity we may take the midpoint of this interval, or

$$f_c = \frac{0.771 - 0.752}{2} + 0.752 = 0.762$$

The value of f_c determines the *intercept* of the discriminant boundary line in Figure 10; the ratio $\frac{c_x}{c_y}$ determines its *slope*.

In forecasting the sale of canine coats in Kansas City (denoted by the * in Figure 10), we first compute its f value (which from the data in

Table 7 is calculated at 0.778), and then compare it with f_c . Since $0.778 > 0.762$ we conclude that Kansas City is most likely to be a "good" sales producer. Since 0.778 is within the area of overlap, however, we must recognize that our forecast is subject to a substantial amount of uncertainty.

Assumptions and Interpretations. The values of c_X and c_Z can be interpreted in terms of the relative influence of X and Z on sales; in the same manner as for partial regression coefficients. Since the zero point on the f scale is arbitrary, however, only the ratio $\frac{c_X}{c_Y}$ is meaningful; their absolute values have no significance.

While the assumptions necessary for the attainment of unbiased estimates of c_X and c_Y have not been worked out formally, as was the case for regression, it seems obvious that something analogous to regression assumption (4) is required if the parameters c_X and c_Y are to be interpreted correctly.

Variables that contribute to sales success but are left out of the discriminant function must be uncorrelated with X and Z if the estimates of c_X and c_Z are to be unbiased.

Again following the regression pattern, we can say:

The presence of excluded variables which are correlated with X and Z will not affect the efficiency of forecasts of sales success for new areas, as long as the offending correlations remain constant.

Note also that predictions based on the value of only one of the independent variables, X or Z , will be much less efficient than those based on both of them together. Similarly, the efficiency of the multiple discriminant function given above may be improved by the addition of new independent variables. A multiple discriminant function involving X , Z , and some other variable might be able to reduce the region of uncertainty, or even discriminate exactly between sales success and failure.

FACTOR ANALYSIS

Like all the other statistical procedures discussed in this paper, factor analysis is a device for reducing an extensive body of data into a more compact form.²³ We desire to find the set of factors, or *principle components*,²⁴ which can effectively summarize the information contained in the sample. The user of factor analysis focuses on the set of variables for which information has been collected and poses the following question:

²³ Harry Harmon, *Modern Factor Analysis* (Chicago: University of Chicago Press, 1960).

²⁴ There is a technical difference between "factor analysis" and "principle components analysis." In the former, estimates of the amount of relationship between each variable and all of the others are substituted for the "ones" in the diagonal of the correlation matrix before analysis. These estimates are called communalities (see Harmon, *ibid.*). The discussion in this paper will be based upon principle components analysis.

"Can the information contained in the original variables be summarized in a smaller number of new variables?" In terms of our example, we wish to find one or two new variables that will summarize the original three: sales (Y), income (X), and temperature (Z).

Like all of the other multivariate techniques discussed in this chapter, that of factor analysis is based upon the information contained in the matrix of correlation coefficients. Since correlations are measures of linear association only, factor analysis is capable of summarizing only linear relationships; the new variables (factors) will be related to the original ones by means of linear functions, as defined below. All of the subsequent discussion presupposes the existence of strictly linear relationships.

The technique of factor analysis was originally developed by psychologists who needed to isolate the components that were common to a large number of interrelated variables. Several hundred questions may appear on a personality test, for example, but it is unlikely that the test really measures several hundred *different* aspects of personality; many of the questions really measure the same thing. By using factor analysis, the analyst can identify the separate aspects of personality and determine the relative amount of information that any given question contributes about each.

Similar problems appear in marketing: consumers may be asked a series of interrelated questions; the same product may be tested in different ways; or various socioeconomic statistics collected. If the original number of variables is too large to deal with directly, or if the variables in certain subgroups are too highly interrelated to allow the use of multiple regression or discriminant analysis (recall the problem of multicollinearity discussed under multiple regression above) a factor analysis may be indicated. Examples of both applications appear in the readings.

The technique can be illustrated within the context of our canine coats example. It begins with the simple correlation matrix, as given in Table 5.

First, let us postulate that there exist two new variables, V and W , that will summarize most of the information in Y , X , and Z . They are defined as follows for each observation ($i = 1, \dots, n$) in the sample:

$$\begin{aligned} Y_i &= c_{YV}V_i + c_{YW}W_i + u_Y \\ X_i &= c_{XV}V_i + c_{XW}W_i + u_X \\ Z_i &= c_{ZV}V_i + c_{ZW}W_i + u_Z \end{aligned}$$

The u 's represent the difference between the summarized and actual values of X , Y , and Z .

In order to maximize the total amount of information contributed by the factors V and W , we often require them to be uncorrelated, that is:²⁵

$$r_{VW} = 0$$

²⁵ The condition that $r_{VW} = 0$ (which is known as the *orthogonality* requirement) is not a necessary one. *Oblique* factor analyses, where r_{VW} is determined by the statistical fit and is not necessarily equal to zero, are also possible (see Harmon,

The problem is to find values for the c 's, which are called *factor loadings*, that are consistent with zero correlation between the factors (V and W) and result in the smallest total sum of squares for u_Y , u_X , and u_Z .

Note that the values for the V_i and W_i are not known. If they were, the c 's could be easily determined by separate multiple regressions of Y , X , and Z on V and W . Since each multiple regression requires that its Σu^2 will be minimum, the sum of Σu_Y^2 , Σu_X^2 , and Σu_Z^2 is sure to be at a minimum also. *Unlike multiple regression, the method of factor analysis produces estimates for all the c 's without ever needing (or finding) values for the V_i and W_i .*

The factor analysis proceeds in two distinct stages:

1. Preliminary estimates of the factor loadings are obtained, together with an assessment of the relative amount of information about each variable (X , Y , and Z) that is contributed by V and W . Unfortunately, there are a great many sets of c 's that satisfy the minimization criteria of factor analysis given above (strictly speaking, there are an infinite number); the computational procedure used to generate the preliminary estimates picks out one particular set in a rather arbitrary manner. Therefore, there is no guarantee that the results from phase (1) will be particularly meaningful in the context of the analyst's problem.

2. The preliminary estimates are translated into a form that (a) still meets the criteria given above, and (b) is subjectively meaningful to the analyst. The method of transformation is that of *rotation of axes*, which is discussed in most elementary analytic geometry texts;²⁶ all sets of loadings which differ only by rotation of axes are equivalent as far as the criteria of factor analysis are concerned.

Interpretation of the factor loadings can best be made if we keep the following fact in mind:²⁷

The square of any factor loading c_{ij} is equal to the proportion of the variance of variable i which is explained by factor j .

For example, c_{YV} (rotated) explains $(0.959)^2$ or about 92 percent of the variation in sales.

Table 9 gives the preliminary and rotated estimates, and contributions to variance, for the factor loadings computed from Table 5. Refer to Part C: the squares of c_{YV} and c_{XV} are 0.919 and 0.822, respectively, allowing us to associate principle component V with sales (Y) and income (X). Factor V explains only 7 percent of the variance of Z , but W explains almost all of Z and very little of Y and X : therefore W is associated mainly with Z .

Looking at the relationships between Y , X , and Z , we see that Y and X are both associated with the same factor, and therefore must be associated (*ibid.*). There is some evidence that oblique solutions may be particularly useful in marketing applications.

²⁶ Cf., George B. Thomas, Jr., *Calculus and Analytic Geometry* (Cambridge, Mass.: Addison-Wesley Publishing Co., Inc., 1954) p. 233.

²⁷ Tintner, *op. cit.*, p. 112.

TABLE 9
FACTOR LOADINGS FOR CANINE COAT MARKET
DATA*

A. Preliminary estimates:

	<i>V</i>	<i>W</i>
<i>Sales Y</i>	c'_{YV} 0.914	c'_{YW} 0.349
<i>Income X</i>	c'_{XV} 0.955	c'_{XW} 0.187
<i>Temperature Z</i>	c'_{ZV} -0.739	c'_{ZW} +0.672

B. Rotated estimates:†

	<i>V</i>	<i>W</i>
<i>Sales Y</i>	c_{YV} +0.959	c_{YW} -0.195
<i>Income X</i>	c_{XV} +0.906	c_{XW} -0.353
<i>Temperature Z</i>	c_{ZV} -0.264	c_{ZW} +0.964

C. Contributions to variance (squares) of rotated estimates of factor loadings:

	<i>V</i>	<i>W</i>	<i>Total</i>
<i>Sales Y</i>	c_{YV}^2 0.919	c_{YW}^2 0.038	0.957
<i>Income X</i>	c_{XV}^2 0.822	c_{XW}^2 0.125	0.947
<i>Temperature Z</i>	c_{ZV}^2 0.070	c_{ZW}^2 0.928	0.998
<i>Total</i>	1.811	1.091	2.902
<i>Proportion of total variance: (Total/n)</i>	0.604	0.364	0.967

* Based on the simple correlations in Table 5.

† Rotation performed by the *Varimax* method, Henry F. Kaiser, "The Varimax Criterion for Analytic Rotation in Factor Analysis," *Psychometrika*, Vol. XXIII, No. 3 (1958), pp. 187-200.

ciated with each other. Only Z is related to the second factor, which indicates that Z is not strongly associated with either Y or X.
The proportion of the total variance of Y, X, and Z that is explained by each factor is equal to the total of its contribution to that of each

variable separately, divided by the number of variables, as shown in the bottom line of Table 9-C. Factor *V* accounts for 60.4 percent and factor *W* for 36.4 percent of the total amount of information in the sample. Likewise, the sum across each row in Table 9-C is the proportion of the variance of each variable that is explained by all the factors taken together: for example, the two factors account for 96 percent of the variance of *Y*, 95 percent of *X*, and almost 100 percent of *Z*. They explain 96.7 percent of the total variation of all the variables.

The interpretation of factor loadings in terms of structural relationships between the variables requires that excluded variables be uncorrelated with the ones in the analysis. If income (*X*) had been neglected, for example, the loading of *Z* upon *V* would have been large and negative because of its negative correlation with *X*. This would lead us to say that lower winter temperatures are associated with more sales, a conclusion we have already found to be erroneous.

The rotated factor loadings for all three variables are in general agreement with the results obtained by multiple regression above, except that the sign of the dependence of *Y* upon *Z* has been reversed. Since the amount of relationship is shown to be small, the error does not lead to a serious misinterpretation within the factor analysis framework. It does serve to indicate that factor analysis is a less efficient procedure than is multiple regression. The latter should be used whenever the dependent variable can be clearly specified, and the independent variables are relatively few in number (usually less than 10 or 15) and not highly correlated with one another; the canine coats data is best analyzed by means of regression. Where these conditions are not met, factor analysis can be a powerful technique. Its use in marketing analysis should be encouraged.

SUMMARY

Five methods of summarizing data for interpretation have been discussed in this paper. Each is appropriate for dealing with particular kinds and sizes of problems.

1. Cross-classification: for use where the extent and nature of relationships between variables are to be evaluated graphically, or directly from tables.
2. Correlation (simple, partial, and multiple): for use where an objective summary of the *extent* of the linear relationship between two or more variables is desired. Numerical data must be available for all variables.
3. Regression (univariate and multiple): for use where summary measures of the *extent* and *nature* of the linear relationship between a dependent and one or more explanatory variables are required. These may be needed for forecasting or for identifying the effects of two particular variables. Numerical data are required.
4. Discriminant analysis: for use where the dependent variable is measured in dichotomous terms: for example, "yes" or "no" or "good" or "bad." The objectives are the same as those for regression.
5. Factor analysis: for use where the information in one set of variables must be summarized in terms of another, smaller set, that is linearly related to the original set. All data must be measured numerically.

The assumptions and conditions required in order for each of these techniques to be used correctly were considered. The most important requirement for isolating the effects of different variables was that all significant but neglected variables be uncorrelated with the explanatory variables used in the analysis; this condition was crucial no matter what statistical method was adopted. The requirements for obtaining valid forecasts are much less stringent: the correlation between excluded and explanatory variables must remain constant over the forecast period, but need not be zero. Other important assumptions were considered explicitly in the discussion of regression analysis.

Finally, the analyst must be fully familiar with both his problem and data. He must be prepared to make subjective judgments based on his prior knowledge, regardless of the technique utilized.

SUGGESTED READINGS

- ANDERSON, R. L., and BANCROFT, T. A. *Statistical Theory in Research*. New York: McGraw-Hill Book Co., Inc., 1952.
- DIXON, WILFRED J., and MASSEY, F. J., JR. *Introduction to Statistical Analysis*. New York: McGraw-Hill Book Co., Inc., 1951.
- EZEKIEL, MORDECAI, and FOX, KARL A. *Methods of Correlation and Regression Analyses*. New York: John Wiley & Sons, Inc., 1959.
- HARMON, HARRY. *Modern Factor Analysis*. Chicago: University of Chicago Press, 1960.
- KEMENY, JOHN G.; MIRKIL, HAZLETON; SNELL, J. LAURIE; and THOMPSON, GERALD L. *Finite Mathematical Structures*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1959.
- KLEIN, LAWRENCE R. *An Introduction to Econometrics*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1962.
- SCHLAIFER, ROBERT. *Probability and Statistics for Business Decisions*. New York: McGraw-Hill Book Co., Inc., 1959.
- SNEDECOR, GEORGE W. *Statistical Methods*. 5th ed. Ames, Iowa: The Iowa State University Press, 1956.
- TINTNER, GERHARD. *Econometrics*. New York: John Wiley & Sons, Inc., 1952.
- VALAVANIS, STEFAN. *Econometrics—An Introduction to Maximum Likelihood Methods*. New York: McGraw-Hill Book Co., Inc., 1959.
- WOLD, HERMAN. *Demand Analysis*. New York: John Wiley & Sons, Inc., 1953.

Complex Interactive Models

ALFRED A. KUEHN

THE PREVIOUS TWO CHAPTERS IN THIS SECTION ON FUNDAMENTALS DISCUSSED techniques of research design and statistical analysis directed at developing an understanding of the relationships between marketing variables. In this chapter we direct our attention to the construction and analysis of models that are more complex than those normally encountered in statistical research. In the last few years the use of models in the solution of complex business problems has grown at a rapid rate. The impact of model building upon theory and practice in marketing is becoming widely recognized.

The following subjects are treated in this chapter: (1) the concept of a model and the development of model building; (2) the types of models being applied to marketing problems; (3) the use of simulation to study the implications of interactive models; (4) the application of heuristics to the solution of complex marketing problems; and (5) the extension of simulation to operational gaming. Points (3), (4), and (5) will be discussed in some detail because of the promise they show for the development of marketing theory and practical techniques of marketing analysis.

THE CONCEPT OF A MODEL

Webster defines the term "model" as a pattern of something to be made, a style of structure, an archetype, a representation of a thing. The use of models in the design of architecture, aircraft and automobiles, furniture and clothing is familiar to most of us. Working models are used by engineers to provide estimates of the performance of the devices the models represent. In each instance the model gives new perspective to its creator and enables him to communicate his thoughts to others more effectively. The problems with which the designer is attempting to cope in the above examples are too complex for solution without the perspective gained from the use of models. In most cases the concepts developed are also of such a nature that they cannot be communicated adequately by resorting to verbal description.

Models are useful only insofar as they adequately represent the essence of the problem studied. A model should be as simple as possible, yet incorporate all of the important aspects of a problem. Models can be misleading—the results provided are no better than the model from which they are derived. Thus model building is not a simple art but rather requires that painstaking attention be given to developing an understanding of the process or system being represented. The ultimate test of the model's adequacy is whether the model provides the perspective needed to reach improved decisions or new understanding.

It is not always convenient or even possible to build physical models of the processes we seek to study. There is, for example, no physical analogy with which to build a working model descriptive of the influence of advertising upon the subsequent purchasing behavior of consumers. Nevertheless, it may be possible to construct mathematical or other kinds of abstract models to describe the effects of these influence processes. To the extent that the essence of the advertising problem can be captured by a model, we have a powerful tool with which to test and extend advertising theory. The approach is to first build as realistic a model as possible, then to study its performance and seek to understand the factors leading to its deviation from reality, and finally to repeat the cycle in an attempt to correct the deficiencies observed. This process forces one to give explicit definition to hypothesized relationships, places already recognized facts in perspective, and points out unforeseen implications of the theory (model) being constructed. In the long run it results in the cumulative development of theory.

A model is a simplified representation of a concept, system, or process usually expressed as a mathematical or logical relationship. The first models developed in physics employed nothing more complex than algebraic equations. The laws of levers expounded by Archimedes, the law of the pendulum discovered by Galileo, and the law of gravitation formulated by Issac Newton are examples of the early use of such models. More recently, the development of the calculus and statistical theory and the advent of electronic computers have broadened the horizon for quantitative application. In the analysis of nuclear reactions, for example, it is now possible to study in probabilistic terms the movement, collision, scattering, and disintegration of individual particles of matter. Insofar as these models adequately simulate the behavior of the reactor system, design engineers may use the results obtained therefrom to reduce or eliminate the risks and costs associated with the development of physical "working models."

Economics was the first of the social sciences to resort to model building in pursuing the construction and communication of theory. As in the case of physics, the development of model building in economics was closely related to the availability and recognition of mathematical techniques that are useful in representing processes and systems of

theoretical interest. As a result of advances in probability theory, decision rules based upon perfect forecasts of the outcome of events have been replaced or supplemented by decision rules which outline optimal behavior in the face of uncertainty. The availability of electronic computers has also opened new doors by enabling economists to work with mathematical relationships (models) of other than classical form. These complex relationships are studied using the technique of analysis known as "simulation," which will be discussed later in this chapter.

CLASSIFICATION OF MODELS

Models have been categorized in several ways: static or dynamic, deterministic or probabilistic, empirical or theoretical, and normative or descriptive. In addition, the language in which models are expressed may be that of mathematical relationships or logical statements such as "and," "or," "if," and "then." Virtually all of the models that have been constructed in marketing have been static and deterministic. Some have been primarily theoretical in character, others largely empirically derived; very few have foundation in both theory and empirical evidence. Both normative and predictive models are common in the literature. While most models have made use of the language of mathematics, examples of the use of logical statements are beginning to appear.

Static models differ from their dynamic counterparts in that the former do not explicitly take time into account as a variable. The static model is of primary interest in the analysis of equilibrium conditions—for example, the state of the market that would eventually be attained if all relevant variables were to remain constant over time. In marketing applications, such a model would be of most interest when there is a very short period of adjustment on the part of consumers, distributors, and/or manufacturers to a change in market conditions.

Dynamic models are designed to describe conditions both at equilibrium and during the period of market adjustment (transition period), and are for this reason more difficult to construct. Not only must the model correctly depict the equilibrium relationships between market variables, but it must also be capable of representing the process of adjustment. The dynamic model has numerous advantages with respect to its static counterpart. For example, let us assume that a change in shelf display is made for a single brand of product in a supermarket. If the effect of this alteration in shelf display is an immediate and lasting change in the brand's share of market within the store, a static model provides a satisfactory description of the situation. If, on the other hand, customers adjust slowly to the new display, the static model is at best an incomplete representation of the consumer's reaction to the new set of market variables. At worst, the use of a static model in such a situation leads to erroneous results; researchers using static models of the "before-after" type must be careful not to accept empirical data obtained during

the transition period as being a measure of equilibrium conditions. The availability of a sound dynamic model to describe the process of consumer adjustment would avoid this source of error, enable researchers to evaluate the overall effect of the display change more quickly, and provide better information to management for decision-making purposes.

Deterministic models assume an exact relationship between sets of variables; probabilistic or stochastic models couch relationships in probability terms. The deterministic model is less demanding computationally and as a result has tended to be more widely used. It seems likely that the bulk of model building in marketing in the foreseeable future will continue to rely upon such models. On the other hand, probabilistic models do appear to offer definite advantages in some areas of application. With the advent of computers, they become operationally feasible. The paper by Orcutt reprinted elsewhere in this book provides an excellent example of the use of stochastic models in simulation.

Models have been classified as empirical or theoretical on the basis of the evidence that supports their use in business applications or as a starting point for subsequent research. It is, of course, desirable that a model have foundation in both theory and experimental evidence.

Empirical models are developed with reference to observed real-world relationships. In the construction of an empirical model the researcher crystallizes and organizes his perception of the structure of the system being studied in such a manner that it can be tested and communicated to others. The sequence of steps involved in the construction and testing of an empirical model is outlined in Figure 1.

Theoretical models take as their starting point abstract principles and theory. Such an approach to model construction is useful because it (1) forces the researcher to specify explicitly all of the elements of his theory, and (2) permits the tracing of its implications. Figure 2 outlines the process by which a theoretical model is built and evaluated.

Models may also be classified as normative or descriptive. A normative model is intended to be a guide to the operation of a business or other economic unit. Given the goals of a firm, such a model would prescribe

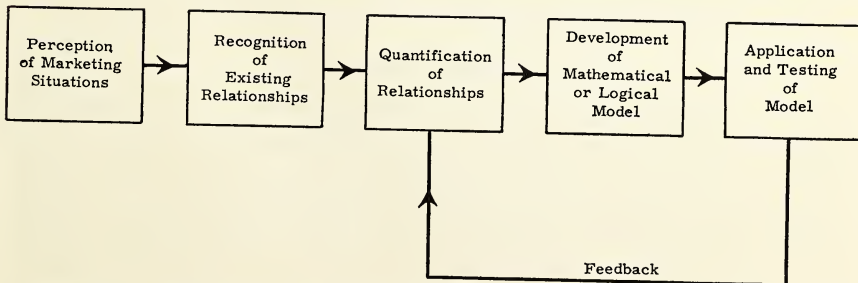


FIGURE 1. Model building by induction from empirical observation. Adapted from William Lazer, "The Role of Models in Marketing," *Journal of Marketing*, Vol. XXVI, No. 2 (April, 1962), p. 10.

desirable rules of operation. The heuristic programming procedures discussed later in this chapter provide an example of the normative approach. Descriptive models, on the other hand, are useful for illustrating the actual behavior of a system, and in some cases for predicting its future course. Most of the simulations discussed in the marketing literature are of the descriptive type.

The advent of the computer has had a dual impact upon the construction and analysis of models. First, the vast increase in computational speed and accuracy makes feasible the use of previously known statistical techniques, such as multiple regression and factor analysis, on a scale never before attempted. Thus, statistical estimation procedures can be applied to a wider range of empirically based models. Second, the computer

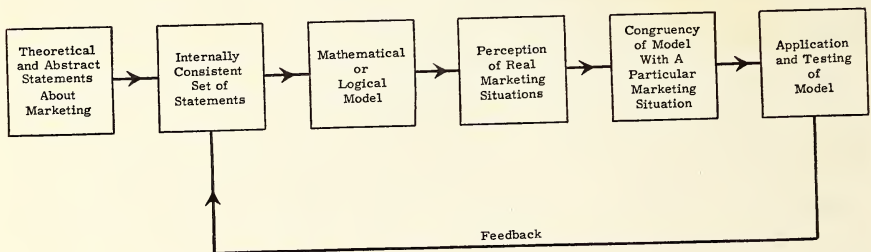


FIGURE 2. Model building by deduction from theoretical principles. Adapted from William Lazer, "The Role of Models in Marketing," *Journal of Marketing*, Vol. XXVI, No. 2 (April, 1962), p. 12.

can be used to trace the implications of any particular model: a "solution" is obtained by programming the computer to calculate answers to the mathematical equations or to apply the logical rules embodied in the model. While many of the methods involved have been known for some time (for example, the numerical solution of differential equations, and Monte Carlo analysis), the availability of computers has led not only to expanded application, but also motivated researchers to extend and improve these methods of analysis. These techniques served as a basis for the early simulation models, and continue to provide important elements for many of the research applications broadly referred to as simulations.

SIMULATION

The term "simulation" has come to be applied to the "process of conducting experiments on a model instead of attempting the experiments with the real system."¹ Wind tunnel and towing tank tests, for example, are physical experiments that make use of scale models of the relevant features of aircraft and marine hulls, respectively. Most business simulations are couched in terms of computational rather than physical experi-

¹ Jay W. Forrester, *Industrial Dynamics* (Cambridge, Mass.: The M.I.T. Press; and New York: John Wiley & Sons, Inc., 1961), p. 18.

ments; a mathematical or logical model is translated into the input language of a digital computer, which then performs the arithmetic or logical operations necessary to obtain a series of results from the model.

Simulation appears to mean widely different things to different individuals. Perhaps because of the current popularity of the term, researchers today tend to title their work as simulation studies whereas five to 10 years ago it was the rage to be doing motivation research. Only with the passing of time will we know what the term "simulation" has come to mean and whether additional terminology will have to be coined to differentiate between the various types of studies now being called simulations.

In an attempt to define simulation as applied to marketing in a more meaningful way, Kuehn and Day have restricted the use of the term to the tracing out of the implications of models of some phase of business activity, which possess all of the following characteristics:

1. The structure of the model seeks to represent the operational characteristics of some segment of the actual business world, be it a part of a firm, an entire firm, an industry, or even the whole economy.

2. The model containing the essential features of the simulated business operation is an abstract, computational system of mathematical and logical expressions designed to accept inputs depicting actual or hypothesized changes in the real situation and to provide outputs approximating the operating results of its real world counterpart.

3. The model is a self-contained system with a structure (consisting of functional relationships and parameter values) that remains stable over time. The structure is not altered by changes in the environment.²

While the above definition of simulation is restrictive with respect to the use of models, it is not restrictive in terms of their structural details. A simulation can be based on a stochastic model in which the structural relations are used to generate probabilistic predictions of outcomes in a manner similar to Monte Carlo analyses, or it can consist entirely of deterministic relationships. The model could contain a sequence of logical operations or consist of a purely mathematical structure. Most large simulation models are likely to contain elements of all of these structural forms.

To the extent that they deal with relevant business systems, all of the types of models discussed in the preceding section of this chapter may properly form the basis for a simulation study. While we might wish to exclude the use of a computer to perform routine clerical work or to solve complex business problems by means of the complete enumeration of alternatives from the definition of simulation, the dividing line between such applications and the description of individual and organizational problem solving may often be difficult to determine. After all,

² Alfred A. Kuehn and Ralph L. Day, "Simulation and Operational Gaming," in Wroe Alderson and Stanley J. Shapiro (eds.), *Marketing and the Computer* (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1962).

even a computer program prepared to perform routine clerical tasks might be regarded as a simulation of a clerk insofar as the clerk might otherwise have performed those functions in essentially the same step-by-step manner. Therefore, it would appear to be appropriate to conclude that the *purpose* of the analysis should determine whether or not it may properly be called a simulation. Only if the goal is to describe, understand, or improve the process, or to predict its future course, would the analysis be a simulation. Thus routine data processing would be excluded from the definition.

One other definitional point should be noted, namely, the difference between static and dynamic simulations. The distinction between static and dynamic models, discussed in the preceding section, is particularly important within the context of simulation experiments. While the term "simulation" was originally applied only to dynamic studies, in the last few years a number of static studies have also been referred to as simulations. The Shycon and Maffei warehouse location simulation is a case in point.³ A model of the firm's distribution system was prepared, including the location of customers, their annual product mix of orders, estimates of transportation costs from point to point, and warehouse operation and inventory carrying costs computed as a function of the annual volume of goods moving through the warehouse. An experiment consisted of choosing an arbitrary set of warehouse locations and calculating the total cost of distribution implied by the model. Since the relations embodied in the model might be considered as an "average" state of the distribution system (that is, time was not included in the model as a variable), the simulation is a static rather than a dynamic study of warehouse operations.

According to the definition given above, every experiment performed upon a business-oriented model can be classed as a simulation. In this sense, even trial and error calculations performed to estimate break-even points under various simple pricing and cost assumptions would properly be classified as a simulation study. In effect, static simulations of the Shycon-Maffei type and trial and error break-even analyses differ only in terms of the number of calculations to be made, a difference which may, however, be staggering from a practical point of view. Dynamic simulations are qualitatively different from their static counterparts and offer a much richer field for analysis. This difference arises because of the inclusion of the time variable which permits analysis of systems in disequilibrium, the effects of cycles and trends, and the impact of feedback upon behavior. The discussion below will be restricted to dynamic simulations.

Dynamic models which are designed for simulation studies usually contain some or all of the following elements: (1) states, (2) response functions or operating characteristics, (3) feedback mechanisms, and

³ Harvey M. Shycon and Richard B. Maffei, "Simulation—Tool for Better Distribution," *Harvard Business Review*, November-December, 1960.

(4) exogenous or independent variables. Each of these will be discussed in turn.

A dynamic model may be viewed as a theory which attempts to relate the condition of a system at one point in time to conditions at one or more previous times. An extremely simplified model might take the form:



The state of the system in time interval t "causes" a response that leads, in turn, to a new state in the subsequent interval $t + 1$. The new state then causes further responses, and the process continues. This model might also be represented in terms of the functional notation

$$S_{t+1} = f(S_t)$$

This is read as follows: "the state of the system at time $t + 1$ is a function of (that is, depends upon) the state of the system at time t ." The notion of causality implies that the function is not reversible. It would not make sense to assert that S_t depends upon the subsequent state S_{t+1} . The states S_t and S_{t+1} might be called the input and output states, respectively, for the response function.

In the context of the simplified model presented above, the term "state" refers to the current set of values for all of the variables included in the system. That is, the level of inventories, the volume of sales, and the degree of confidence that businessmen feel toward the economy might all be considered as aspects of the state of a system.

The response function lies at the very heart of the dynamic behavior of the model. Depending on the context, it might refer to the decision-making characteristics of individuals or organizations, the behavior of aggregates of these units as in national income and other macroeconomic studies, or the characteristics of natural or technical phenomena. The terms "operating characteristic," "decision function," and "behavioral relation" have all been used as synonyms of the term "response function."

While the state of the system can in principle be measured directly in an empirical study of the system being modeled (such measurement may be prohibitively expensive in practice), the response functions must usually be inferred by inductive means. They may be estimated from empirical data by using the techniques discussed in the previous chapters on experimentation and statistical analysis of relations between variables.

For certain purposes, it may not be necessary to obtain empirical estimates of response functions. For example, insofar as a researcher is primarily interested in studying the sensitivity of a system to individual parameters of the response function, the simulation model may be evaluated with the parameters set at various "plausible" values.

The simple model given above may be complicated in a number of different ways. First, a multitude of causes can operate to induce any given response. Second, there is no need to restrict inputs to the state of the system in the time interval immediately preceding the response; the existence of lagged effects is commonly recognized. Third, feedback loops may be present. The term "information feedback system" has been used by Forrester to describe a situation where

the environment [state] leads to a decision that results in action which affects the environment and thereby influences future decisions.⁴

In the notation of this chapter, feedback occurs wherever a response function takes for its input a portion of its own output from a previous period. When lagged effects are introduced into a feedback network, the system can exhibit a tendency to amplify fluctuations in input variables; such models may exhibit markedly unstable dynamic characteristics. On the other hand, the presence of feedback loops can tend to mask inconsistencies between the behavior of parts of the model and its real-world counterpart. The feedback network may operate as an automatic control system, tending to keep the output within plausible bounds even though some of the response functions in the model may be grossly unrealistic.

Exogenous or independent variables serve as inputs but do not appear as outputs for any response function. They may also be referred to as environmental variables. The decision as to whether a given variable should be treated as part of any self-contained dynamic system, or considered to be exogenous, depends upon the purposes of the research. Personal income might well be exogenous for a simulation of the distribution-demand characteristics of television receivers, for example, but should definitely be treated as a dependent variable in experiments with a macroeconomic business cycle model.

Finally, the model must be structurally stable over time, since it represents a theory which must be tested against the observed behavior of the actual system on which it is based. If the model's results are not consistent with the behavior observed in the actual situation, structural adjustments in the model are required. Such changes amount to revisions in the theory of the behavior of the actual system to which it corresponds.

HEURISTIC PROGRAMING

Simulation, as defined above, represents an approach to describing, studying, and analyzing the systems which underlie the marketing process. Once a simulation model has been constructed and tested, it may be applied in practice to predict the effects of alternative marketing decisions, policies, and strategies. In this way, simulation can guide

⁴ Forrester, *op. cit.*, p. 14.

management's thinking and decision making with respect to the solution of complex marketing problems.

Heuristic programing is more directly oriented toward the solution of specific problems. Heuristic programs are computational procedures designed to produce "acceptable" solutions to problems which have been cast into the framework of mathematical or logical models. In the simplest form of this approach, one might take the tactics or "rules of thumb" used by a person in solving everyday problems and program them for use on a digital computer. The combination of human cunning and the large memory and high speed of the computer provide a methodology for problem solving not heretofore available. In effect, the computer makes possible the solution of large-scale problems which man alone could not solve (using a similar approach) because of the inordinate amount of time that would be required.

Newell, Shaw, and Simon have used the term "heuristic" to "denote any principle or device that contributes to a reduction in the average search for the solution" of a problem.⁵ Any rule or computational procedure which restricts the number of alternative solutions in the analysis of a problem might then be considered to be a heuristic. It is not required that an individual has actually used such a rule in the past. New heuristics can be developed to take advantage of the capabilities of computers. The heuristic approach is, however, based upon the human trial and error process of reaching "acceptable" solutions to complex problems for which we do not expect, at least in the short run, to be able to devise a "best possible" solution.

Operations research as applied to business problems (that is, management science) has, for the most part, centered about analytic models specified in such a way that optimum solutions can be obtained by the use of specific computational routines. The application of such computational procedures, generally referred to as algorithms, might be thought of as being technique oriented. That is, in constructing models, the emphasis is upon describing the process being studied in an analytical framework which permits the computation of the best possible solution (or solutions) to the problem as it is described within the model. The requirements of the existing algorithms (for example, the linearity assumption in linear programing) tend to influence the researcher to model the problem in such a way that the assumptions required by the computational algorithm are not violated. As a result, the question of how well the model captures the essence of the system being studied can inadvertently receive only secondary consideration.

In contrast, the heuristic programing approach generally permits much greater flexibility in the modeling of the problem. Without the

⁵ A. Newell, J. C. Shaw, and H. A. Simon, "The Processes of Creative Thinking," The RAND Corporation Paper, P-1320, August, 1958.

constraints imposed by rigid computational algorithms, the researcher is free to construct more realistic models of the problem for which a solution is desired. This advantage in the modeling of the problem must be balanced against the advantage of optimizing algorithms, namely, that once a solution is obtained by use of the latter procedures it is guaranteed to be optimal with respect to the model of the problem. There is generally some uncertainty as to how good a heuristic solution is with respect to a specific problem. There is no guarantee of an optimal or even a near-optimal solution inherent in the heuristic approach, since the

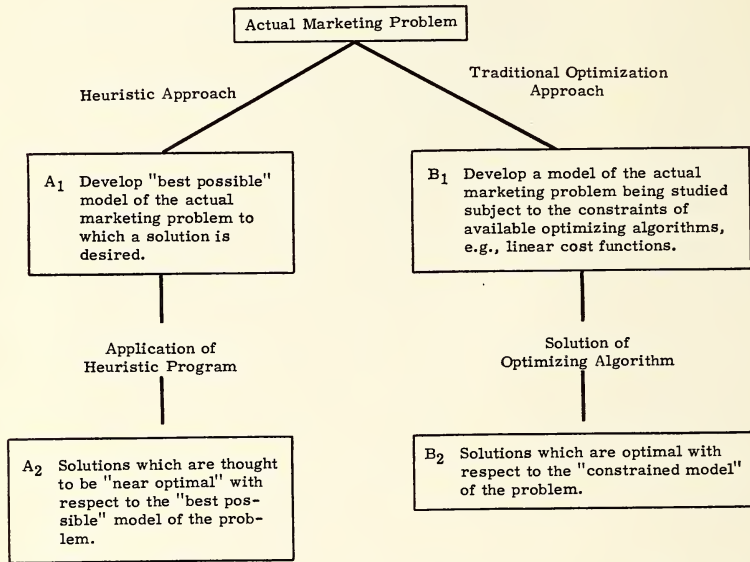


FIGURE 3. Comparison of heuristic and algorithmic approaches to problem solving.

heuristic approach is basically an inductive process. Confidence can be acquired through repeated application of a heuristic program to a variety of related problems, particularly insofar as these solutions can be compared to solutions derived by alternative methods. Nevertheless, the certainty that is associated with a deductive algorithm is necessarily lacking.

Figure 3 illustrates the heuristic approach in comparison with the traditional optimizing algorithmic approach to model building. The appeal of the latter approach has been its guarantee of optimal solutions (B_2) to the problem as described in the constrained model (B_1). Insofar as the model is a good representation of the actual marketing problem, and a solution can be reached within practical limits on computational time, this approach is clearly to be preferred to the heuristic approach. However, there is a tendency to use this approach even though lack of flexibility available in the modeling of the problem may not permit an

adequate description of the real-world system being studied. The availability of convenient computational techniques motivates the analyst to cast the problem into the framework required by the optimizing algorithm. The less the analyst knows about the real problem, the more likely he is to believe that the constraints imposed by the optimizing algorithms are not inconsistent with the situation being studied. In such cases the optimum solution to the constrained model (B_1) may be a very poor solution to the actual problem to which a solution is desired.

In the construction of a heuristic program, emphasis is upon developing logical or computational routines which reduce the total problem to manageable size for purposes of analysis. These routines or heuristics should be chosen so as to have high selectivity in eliminating sets of alternatives likely to result in poor solutions while retaining sets of alternatives which have a high probability of yielding optimal or near-optimal solutions. These routines should also be designed with a view to keeping computation time requirements as low as possible. The computational costs of increased complexity in "search" within these routines must be evaluated relative to the increased selectivity likely to result. The analysis of test cases using alternative heuristics may be required to achieve an approximation of the balance desired in this respect.

Perhaps the most disconcerting problem in the construction of heuristic programs for the solution of complex problems is the absence of known "best solutions" for a variety of test cases against which heuristic solutions might be compared. One method of obtaining a partial evaluation of a heuristic program is to apply the program to a model whose optimum can be obtained by use of an optimizing algorithm. In Figure 3, this would be equivalent to setting A_1 identical to B_1 and then comparing the solution A_2 to solution B_2 . Such a comparison can provide some indication of the degree of suboptimality associated with the heuristic program although caution must still be used since the form of the problem in B_1 is not identical to the form of the problem in A_1 . A rough estimate of the magnitude of error that might be introduced into the solution B_2 as a result of the reduced flexibility in modeling the problem at B_1 is possible by first evaluating the operational decision reached in B_2 when applied to the "best possible" model of the problem (A_1), and then comparing the "cost" or "value" of this solution with respect to that computed in A_2 .

The problem of choosing between heuristic programs and optimizing algorithms for the solution of specific problems must be decided on the basis of the error introduced into the operational decisions as a result of the algorithmic constraints in the modeling of the problem at point B_1 versus the suboptimality of the heuristic solution A_2 with respect to the model A_1 . A comparison of the operational decisions reached at A_2 and B_2 , each evaluated with respect to the actual problem would be most desirable. This is not, however, generally possible.

Other approaches to the evaluation of a heuristic program might include the complete enumeration and evaluation of all possible solutions to the "best possible" model (A_1), feasible only for small-scale problems, the testing of heuristic solutions for optimality through the use of appropriate algorithms (possible in certain special cases) and comparison of the heuristic solution with solutions proposed by "experts" in the field. It should be noted that heuristic solutions need not be optimal to be of practical value; all that is required is that they represent some improvement (in terms of the solutions provided and the costs of application) with respect to alternative existing methods in use or available to management.

Clearly, there is a place for traditional optimizing algorithms as well as a place for heuristic problem-solving methods. Such techniques as linear programming, integer programming, quadratic programming, the calculus, and queueing theory are capable of solving a variety of scheduling and resource allocation problems. There are, however, limitations to the complexity and the size of the problems that can be adequately solved by such optimizing methods. In addition, marketing problems frequently cannot be formulated in the framework required by these techniques without making a number of assumptions whose validity is open to question. In such cases, more damage can be done by forcing the modeling of a problem into the required mold than by accepting suboptimal solutions to an improved description (model) of the actual problem. Heuristic programming as a tool for problem solving comes into its own when the assumptions required by rigorous optimizing techniques are either inappropriate to the problem or when, as is frequently the case, the size of the problem prevents the practical application of such techniques.

OPERATIONAL GAMING

The term "operational gaming" was originally coined to mean "the use of war gaming in a context broader than that of military situations alone."⁶ This definition predated the development of the business game, which is generally regarded as the application of war gaming concepts to business situations. The earliest games of this type were frequently called "business war games."

The war game had two basic purposes—training and strategy evaluation. Game structures based on business situations have been used primarily as training devices but they also have potential for use in competitive strategy evaluation and problem solving. Games for business training purposes are now closely associated with the terms "business game," or alternatively, "management game." It seems desirable, therefore, to re-

⁶ Walter E. Cushen, "Operational Gaming in Industry," in Joseph F. McCloskey and John M. Copping (eds.), *Operations Research for Management* (Baltimore: The Johns Hopkins Press, 1956), Vol. XI, p. 358.

serve the term "operational game" to describe game structures specifically designed for use in strategy evaluation and problem solving.

Since operational games are closely related to games used for training purposes, it seems appropriate to review the nature of business games. The general characteristics of the business game can be summarized as follows:

1. The structural aspects of a business situation, real or hypothesized, are depicted in a computational model which accepts inputs reflecting the decisions of participants and provides outputs, such as sales figures, which correspond to the consequences of decisions in the business situation.

2. A number of players, usually organized into teams, represent managers of business units and make the decisions necessary in "running the business." These units (e.g., business firms) compete directly so that the results of one unit's actions are influenced by the concurrent actions of its competitors and a great variety of different solutions can result.

3. The game is a multi-stage process with the stages representing successive time periods. There is a continuity of the basic underlying conditions from period to period but elements of change are introduced as the result of the actions of the players and by changes built into the model, such as seasonal patterns and growth trends.

4. Play of the game is "task oriented" in that players center on one or more major objectives such as the maximization of profit, with performance being judged in relation to the accomplishments of competing teams.⁷

Since business games have been developed as educational tools, the major objective in building the underlying models has been to provide an environment in which learning can take place. The students can thereby gain experience in decision making and a better understanding of the interdependence of the many activities which take place in a business firm. It is believed that the dynamic environment provided by the game model achieves these objectives better than traditional methods of business education such as reading and discussion or the study of cases.

Since the educational objectives of business games are broader than the teaching of the problems and practices of a single industry, the developers of game models have usually used hypothetical industry settings or merely the outward trappings of an actual business situation. Thus business game models typically have not attempted to reflect accurately the complex underlying behavioral mechanisms of the marketplace. However, the basic structure of the business game could provide an interesting approach to the study of alternative strategies for a business firm if the underlying structure actually reflected the major operating characteristics of a specific industry.

By incorporating a comprehensive simulation model of an industry in the game structure, an "operational game" can be developed. The firms

⁷ Adapted from Ralph L. Day and P. John Lymberopoulos, "The Decision Game: Progress and Prospects," *Southwestern Social Science Quarterly*, Vol. XLII (December, 1961), pp. 251-58.

in the game can be established with resources and market positions corresponding to those of firms in the actual industry. Then experienced business executives can play the role of the firms in the industry, making it possible to test proposed strategies in the context of the game. Each strategy could be tested to study its implications and the various ways in which competitors might react to it. If the operational gaming model adequately represents the essential features of the market environment, alternative proposed actions can be tested without incurring the costs and risks involved in actual market tests. However, the techniques of simulation and operational gaming must be improved greatly before results obtained in the artificial environment of an operational game can be expected to closely correspond to real-world results. But even relatively simple models will often reveal relationships and point up opportunities that are overlooked in routine analysis of actual data.

As the number of simulations of actual industries increases, and their quality improves, there will be increasing opportunities for using game models for operational purposes as well as for training.⁸ When more knowledge is gained about the way in which decisions are made within firms, it may be possible to replace the teams of executives who play the role of competitors with computer models. The only human elements in the operational game would then be the executives of the firm using the game.

APPLICATIONS IN MARKETING

Before the development of computer simulation and related techniques, most students of marketing considered the problems in the field to be far too complex for extensive application of mathematical analysis. As operations research gained popularity, its methods were applied to the problems of production management much more rapidly than to the problems of marketing management. A long history of failures of economic and mathematical models to be of much help in directing a firm's pricing, advertising, and new product strategies has made marketing executives wary of quantitative techniques in general. As a result, marketing undoubtedly relies on "intuitive decision making" to a much greater extent than any other area of management, huge marketing research budgets notwithstanding.

One of the major causes of the relative lack of success of quantitative techniques in marketing has been the way in which the field has been partitioned for research. Although marketing analysts have been among the most vociferous critics of classical economic theory and its *ceteris paribus* assumptions, most of the empirical work in marketing has centered on one independent variable at a time. Many attempts have been

⁸ See, for example, Doyle L. Weiss, "Simulation of the Detergent Industry," *Marketing Precision and Executive Action*, American Marketing Association, 1962.

made to relate the effects of advertising to sales volume, or the corresponding effects of price changes, while essentially ignoring other variables and the interrelationships among variables.

Repeated failures of such research to produce clear-cut operational results has led marketing practitioners and researchers alike to look upon marketing as a complex maze of behavioral mechanisms not amenable to mathematical analysis. Consumer behavior has been regarded by many as irrational, inconsistent, and entirely unpredictable. Recent advances in quantitative research methodology directed at bringing order into our understanding of the dynamics of consumer buying behavior have been regarded with skepticism by many marketing executives.

Until the advent of the computer and the rapid development of complex analytical techniques which it has facilitated, models involving large numbers of variables and complex interrelationships were not operationally feasible. Analytic models had to be kept relatively simple. Unfortunately, however, most marketing problems involving the various aspects of competitive strategy appear to be too complex for adequate treatment by simple models. The computer has made it possible to tie together a number of relatively simple analytic models, each dealing with a particular aspect of marketing, into a complex computer model in such a way that the component models interact with one another. Instead of assuming "all other things constant" and looking at only one or at most a few variables at a time, it is possible to study the effects of a large number of variables and complex relationships simultaneously.

It should not be inferred, however, that complex interactive models can be quickly put together and tested as a result of the development of computer simulation techniques. Building a realistic computer simulation of a market is, and probably will continue to be, an extremely involved and painstaking task. An attempt to construct a massive simulation model in one fell swoop is not likely to be successful. Whenever possible, the system to be simulated should be broken down into subsystems for which models can be developed and tested independently. Thus, the construction and testing of a simulation model is much like putting together a picture puzzle. The individual subsystems, each involving a small number of variables, are joined to potentially related parts. The resulting model can then be examined to establish that the appropriate connections have been made. By proceeding in this fashion, each part of the model can eventually be placed in proper alignment with all other parts.

Perhaps the greatest problem in simulation is the one which is frequently sidestepped in current research efforts—the difficulty of testing such models. In the case of a picture puzzle it is generally apparent to all observers whether or not the pieces fit. The evaluation of a simulation is not nearly so simple. A model might be quite inadequate as a simulation and yet appear to the casual observer to be reasonably con-

sistent with reality. Balderston and Hoggatt⁹ have said that "what one typically encounters in a computer simulation . . . is the presentation of a few paragraphs of terse description and a few thousand lines of tangled [computer] coding in one of the many possible coding systems."

It is not easy to evaluate adequately a simulation model. This is particularly the case during the development of the model. Suppose that the time paths of results generated by the model bear little relation to those observed in the real system. Where then is the fault? Are parameters set incorrectly, are minor subroutines in error, or is the entire structure of the model inappropriate to the problem? The overall results from large-scale simulation runs which are not consistent with reality provide few clues to the sources of error.

The computational power of electronic computers enables us to expand greatly the number of variables included in simulation models and to increase the complexity of their interrelationships. The ease with which this can be done gives the researcher a feeling of great power. Unfortunately, means of equivalent simplicity for pinpointing errors in such systems are not available. There is reason to believe that this will continue to be true for a good many years to come. Consequently, it appears necessary that the researcher's attention be directed at partitioning the total problem, analyzing and testing it piece by piece, so that when the segments of the simulation are combined fewer aspects of it will be open to suspicion of error. The basic rule of model building has not been altered by the advent of the computer. That is, concentrate on a relatively few variables at a time at the early stages of model construction. Once the behavior of the simple model is fully understood (that is, "internalized" by the researcher) additional variables can be incorporated. Or a series of small models can be combined into more complex simulation models. The probability of success in attacking a large-scale simulation problem without intermediate testing of submodels is probably no greater than the probability of one's being able to correctly position each piece of a new picture puzzle directly, drawing each piece for placement in a random order.

SUMMARY

Model building, simulation, and the related techniques of heuristic programming and operational gaming represent new analytical approaches for use in the development of marketing theory and applied research methods. There is promise that they will prove to be of real value to marketing management in the solution of a variety of complex problems, including the choice of alternative merchandising strategies.

An attempt has been made in this chapter to point out the distinguishing characteristics of each of the above techniques and to present a balanced view

⁹ F. E. Balderston and A. C. Hoggatt, "The Simulation of Market Processes," Management Science Research Group, University of California (Berkeley), October, 1960.

of the promise and problems associated with their use. The last part of this book presents several examples of simulation studies and outlines an application of heuristic programming. An example of an operational marketing game is not yet available. Researchers are attempting to construct realistic simulations of markets for use in such games but have not reached the stage where a game is truly operational.

SUGGESTED READINGS

- ALDERSON, WROE, and SHAPIRO, STANLEY. *Marketing and the Computer*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1962.
- BASS, FRANK M., et al. (eds.). *Mathematical Models and Methods in Marketing*. Homewood, Ill.: Richard D. Irwin, Inc., 1961.
- COHEN, KALMAN J. *Computer Models of the Shoe, Hide, and Leather Sequence*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1961.
- COHEN, KALMAN J., and RHENMAN, ERIC. "The Role of Management Games in Education and Research," *Management Science*, January, 1961, p. 131.
- FORRESTER, JAY W. *Industrial Dynamics*. Cambridge, Mass.: The M.I.T. Press; and New York: John Wiley & Sons, Inc. 1961.
- MILLER, DAVID W., and STARR, MARTIN K. *Executive Decisions and Operations Research*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1960.
- ORCUTT, GUY H.; GREENBERGER, MARTIN; KORBEL, JANICE; and RIVLIN, ALICE. *Micro-Analysis of Socioeconomic Systems: A Simulation Study*. New York: Harper & Bros., 1961.
- ORCUTT, GUY H.; SHUBIK, MARTIN; CLARKSON, G. P. E.; and SIMON, HERBERT A. "Simulation: A Symposium," *The American Economic Review*, Vol. L (1960), pp. 893-932.
- SIMON, HERBERT A., and NEWELL, A. "Heuristic Problem Solving: The Next Advance in Operations Research," *Operations Research*, Vol. VI (1958), pp. 1-10.
- TONGE, FRED M. *A Heuristic Program for Assembly Line Balancing*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1961.



**READINGS ON
EXPERIMENTATION**

PART
THREE



Laboratory Experimentation

THE READINGS IN PART III ARE DIVIDED into two groups: those dealing with laboratory experiments and those dealing with field experiments. This method of classification was chosen (as opposed to one by marketing problem, by the way in which control groups were used, or by the type of variance reduction technique used) because it seems to come closer than do the other alternatives to achieving two pedagogical objectives:

1. It permits the contrast of experiments within each of the two sections (laboratory and field experiments) in terms of the characteristics of experimental design (for example, how were the controls set up? How were the units to be included in the experiment divided into groups? And so on.).

2. It also permits the contrast of field and laboratory experiments in terms of the extent to which they occur under conditions similar enough to those in practice to allow the results to be used to modify a decision maker's behavior.

There are four readings in each of the two sections. The designs used for laboratory experiments in marketing have tended to be somewhat simpler than those used in the field, and therefore are presented first.

The Alderson and Sessions' "Basic Research Report" tells of a shopping game in which an attempt was made to measure the effect of such variables as number of merchandise lines, depth of assortment, and width of price line on the number of purchases consumers make by store. The game was played by giving each respondent a set of envelopes, each of which represented a given store. Tickets in the envelopes represented particular price lines of given items. The mix of tickets was varied from envelope to envelope so that the respondents reaction to the above variables could be estimated. The technique allows one to use experimentation

to study variables that would be extremely costly to manipulate in the "real" world.

One way of analyzing this experiment is to try to specify the way in which the experimental units were organized in detail, as a basis for making inferences about the effects of number of lines, depth of assortment, and width of price line. Once this has been done, the reader might find it fruitful to ask how the design might be modified so as to allow the extraction of more information from an experiment of this type?

The Gridgemen paper reports an example of one of the most frequent applications of laboratory experimentation: taste testing. Gridgemen describes the results of a taste test which involved determining the effect using glucose as a component in jam on consumer preferences. The results are based on a series of paired comparisons of regular jam (with sucrose) versus jam with glucose added.

Pessemier reports a shopping game which was designed to generate data to serve as the basis for estimating price elasticities by brand for several categories of frequently purchased items, such as cigarettes, toilet soap, and headache remedies.

Respondents were given slips of paper on which were listed the names of several brands in a particular product category, together with their respective prices. The respondent also received a budget of \$1.75 in order to shop the items. The relative price of the particular brand being studied was systematically varied from one trial to another so that information was generated on the relationship between the relative price of the brand and the quantity purchased.

The article entitled "Why Television Commercials Succeed" reports an experimental technique developed by the Schwerin Research Corporation for evaluating television commercials. Respondents were invited to a theater and exposed to a half-hour television show with commercials in their normal spots. Measurements of preference and choice were made before and after the showing. Comparisons of a set of alternative commercials were made by exposing each commercial to a different group of respondents. Of particular interest, from the standpoint of research design, is the use of matching as a device to help insure the comparability of respondents, and the attempts made to check the reliability of the technique.

A brief description of the articles on field experimentation is presented at the beginning of that section.

*Basic Research Report on Consumer Behavior: Report on a Study of Shopping Behavior and Methods for Its Investigation**

.....

ALDERSON and SESSIONS†

ORIENTATION

THIS IS A REPORT ON THE MOST RECENT PHASE OF THE CONTINUING Alderson and Sessions' basic research project investigating consumer motivation. This project is designed to investigate consumer behavior and develop tools of research with which it can better be investigated in the future. Its ultimate object is to provide for ourselves and our clients a better understanding of marketing problems, thereby in turn permitting both them and us to operate more effectively.

Business executives would, of course, like to know in advance how consumers will react should they add a new product line, reduce a price, or sponsor a new television program. Though this sort of information is essential for good decision making, it is rarely, if ever, available.

Worse still, the executive may not even be able to discover the consumer's reaction to his decision afterward. Can any of his sales increase be ascribed to his increased magazine advertising budget, or is it largely the product of rising consumer incomes? When he cuts his price and sales fall, does this mean he has made a mistake, or would fickle customer tastes have led to an even greater decrease in demand if he had maintained his old price? There are always so many things going on at once that it is difficult, and sometimes impossible, to identify the consequences of some particular business decision.

This point is of very considerable importance because it sets a limit to the amount we can learn by experience. If experience cannot be evalu-

* This is an edited version of an original mimeographed report dated April, 1957.

† The firm is now known as Alderson Associates, Inc.

ated, it cannot serve as a reliable guide to future action. As a result the hunches and rules of thumb which guide many advertising, pricing, product line, and other business decisions must be considered suspect.

In these circumstances it is critically important that techniques be developed which permit prediction of the effects of executive action on consumer behavior. One of the primary objectives of the basic research program is the development and testing of just this sort of technique for the acquisition of consumer behavior information.

Behavior Research and Market Experimentation

The term "behavior research" has been chosen to describe an important addition to the tools employed in predicting the effects of marketing decisions on consumer behavior. This approach is characterized by direct experimentation. The consumer clinic conference room is the market researcher's laboratory. There the market situation is duplicated as closely as possible and the response of a sample of consumers to a contemplated business decision is carefully observed. Several groups of consumers may be offered a product at different prices, or they may be asked to read magazines with alternative types of advertisements inserted, and then to make purchase decisions. The subjects can be given money in advance, some of which they may choose to spend on the products in question—this to increase the similarity of their behavior during these games to their real shopping behavior. These, then, are the two cardinal features of behavior research: (a) duplication of the market situation in the laboratory, variation being permitted only in the feature, price, product line, or promotion technique, whose effects are under investigation, and (b) observation of reactions of the experimental subjects.

A clearer understanding of this approach can be had by contrasting it to the two principal alternative methods currently in use: large sample interviews and small sample motivation research studies. The results of large sample interviews can be questioned on several grounds. Consumers may not be candid when answering or they may not know the answer themselves. Few of us can predict how we will behave in a hypothetical situation until we are actually faced with it. Often the mental process that is relevant for the decision is subconscious. In attempting to overcome these problems, researchers have turned to a group of techniques which fall under the label of "motivation research." The primary objective of these techniques (e.g., depth interviewing, sentence completion tests) is to search for the underlying motivations that determine behavior. While motivation research strives to gain insight into the underlying determinants of behavior, its results are often difficult to quantify and interpret. In addition, the caliber of the personnel involved and the length of the interviews frequently make large sample techniques uneconomical.

Two of the goals of the experiments reported in this paper are:

(1) to increase the quantifiability of the results, and (2) to reduce the costs of small sample investigations. If, as preliminary indications suggest, costs can be reduced successfully without substantial reduction in the effectiveness of the methods, the small sample techniques will be enabled to employ larger samples. If numerical measurement of the results is facilitated, more clear-cut evaluation and greater confidence in recommendations will become possible.

It must be made clear, however, that no magic formulas can be expected to achieve these aims. Improvement in the tests in one respect can usually be obtained only at the cost of some weakening in other respects. For example, we are experimenting with Thematic Apperception Tests which are designed to yield fairly comparable and quantifiable data. This can only be done by somehow limiting the rambling in the conversation of the person being tested. In the process he is presumably given less opportunity to let his subconscious show through. Whether the advantages balance off the losses remains to be seen.

More important has been our work on the methods of behavior research. It is believed that this tool can increase the utility and reliability of motivation research. If used with care and common sense, motivation research can be very fruitful.

METHODOLOGY

The data gathering and analysis was organized around a family panel as the primary source of information. Since the household is the central purchasing and consumption unit, it was decided that focus on the family would give us the most useful picture of what goes on in shopping decisions and behavior.

Selection of the Panel

The sample of 61 families was drawn from four tracts in metropolitan Philadelphia. The sample was drawn by selecting tracts whose residents were approximately of median Philadelphia family income as of 1949, and then randomly selecting blocks within these tracts.

The housewives selected for members of the family outlook panel met the following qualifications: (a) they came from an income group which approached the Philadelphia median; (b) they had a friend on the panel (this permitted cross-checking and examination of interactions in shopping behavior); (c) they were between 25 and 50 years of age; (d) they had children; (e) they had lived in the neighborhood at least one year; (f) they owned or rented a home or unfurnished apartment; (g) they were English speaking; (h) they were Caucasian; (i) the husband lived at home.

Because of the controlled heterogeneity of the panel, individual differences in shopping are not ascribable to large differences in socioeconomic status. By choosing the ranges of the variables, however, the sam-

ple was made to be representative of the large middle range of women shoppers. The typical housewife on the family outlook panel was about 36 years old, had one or two children, and her family income was about \$5,000 per year. There were two chances out of three that the family owned its own home and it was probably a row house. The family had lived in the same neighborhood for approximately five or six years, and the husband might have been a craftsman or operative. The likelihood is that both husband and wife graduated from high school.

Interviews

Panel members were interviewed once every two weeks for a total of five interviews. They were asked about shopping trips during the previous two weeks. The respondent was encouraged to talk at length about her shopping experiences and the interviewer directed the conversation so as to obtain factual information pertaining to the items shopped for, items bought, and number and names of stores visited. In general, the interviews revolved about (a) prepurchase matters such as requirements for items desired, prepurchase planning and discussion, influences of advertising and other market information, and the anticipated difficulties, if any, in finding the items; (b) the actual course of the shopping trip—where the shopper went, what she saw, and reasons for purchase or failure to purchase; and (c) postpurchase matters such as use of the items, the degree of postpurchase satisfaction, and rationalization of the decisions as to which items were to be purchased.

In addition to the regular shopping trip information collected at each interview, each panelist was questioned during the original interview about her general shopping habits, preferences on suburban, local, and downtown shopping centers, attitudes toward shopping in general, and major purchases which had been made in the last year.

The purpose of the continuing family outlook panel was to maintain contact long enough to get well acquainted but to stop the interviews before the participants became bored. The interviewer-respondent rapport increased with each visit, permitting the interviewer to pose probing and personal questions which would have been virtually impossible in a single-contact depth interview.

The Shopping Game

In continuing the work started at MIT,¹ shopping games were developed to investigate some aspects of merchandise assortment as they affect store preference. For this purpose three sets of "stores" were designed, each set containing nine stores.

¹ Prior to this experiment, two shopping games had been developed by Wroe Alderson while he was at MIT during the winter of 1953. For a detailed account of these procedures and results, see Alderson and Sessions (now Alderson Associates, Inc.), *Cost and Profit Outlook*, Vol. VII, Nos. 1, 2, and 3 (January, February, and March 1954).

Each "store" was an envelope containing descriptions of 16 items of merchandise clipped from a mail-order catalogue. The first set of nine stores were set up in the manner shown in Table 1.

The store names were chosen as the nine most common surnames in metropolitan Philadelphia, excluding any names of women in the panel. The respondents examined this set of stores at their leisure, with instructions to note the type of each store.

The second set of nine stores had the same characteristics as the first, except that half the items (8 out of 16) in each store were different than they were in the first set. However, all the other characteristics were the same—the store names, the merchandise lines, and the price ranges. This time the women were instructed to look for a "best buy." In real shop-

TABLE 1
CHARACTERISTICS OF THE FIRST SET OF GAME "STORES"

<i>Store Name</i>	<i>Number of Merchandise Lines</i>	<i>Number of Items per Line</i>	<i>Total Number of Items</i>	<i>Price Range (Dollars)</i>
Brown.....	2	8	16	\$6.00- \$ 8.00
Davis.....	2	8	16	5.00- 9.00
Green.....	2	8	16	3.50- 10.50
Hoffman.....	4	4	16	6.00- 8.00
Johnson.....	4	4	16	5.00- 9.00
Miller.....	4	4	16	3.50- 10.50
Smith.....	8	2	16	6.00- 8.00
Williams.....	8	2	16	5.00- 9.00
Wood.....	8	2	16	3.50- 10.50

ping there are significant costs involved in going to additional stores. The baby-sitting bills go higher or the tired shopper grows more exhausted. In the games the women were constrained from going to many stores by the fact that their score diminished by 10 points out of an initial 100 points for each store they examined after the first.

The items were assigned a random value between 751 and 899. The higher the random number the better the buy. Thus the maximum score was 999. In practice, the scoring system went like this. Each time a shopper entered a store (after the first time) it cost 10 points. Once in the store, the shopper looked for the best buy, which was indicated by the highest random number. After looking in the first store, the shopper could either look in another store, which would cost 10 points and might lead to finding a better buy, or she could terminate the shopping trip.

In the third set the stores contained none of the items in either the first or second set. This was explained to the respondents to discourage "item" shopping and to emphasize the necessity of learning the store's character. In the third set there was one winner at each game section and as a prize she was sent the item she had elected to "buy" during the

game. The possibility that she might actually find herself in possession of the item selected served as an inducement to the panelist to take her decision seriously. For the reasons stated in the preceding paragraph the fewer stores a woman visited the more chance she had of winning. This was done by having her forfeit one token out of an initial 10 for each store she visited and then holding a drawing of the remaining tokens for the winner.

The game shows the interaction of the three variables, number of merchandise lines, depth of assortment, and width of price line. In addition it supplies insight into shopping behavior and attitudes.

The game construction is based on the hypothesis that the information shoppers have about the retail market consists of a set of descriptive parameters for each store or shopping area. These descriptions (good style, not too large an assortment, rather high price but worth it) are supplemented by a partial list of expected items available (I know they have Crosley 21-inch TV sets). Since the parametric aspect was emphasized and because of the crowding in time, the stores were described in terms of the lines of merchandise only, rather than on the basis of the actual items offered for sale.

Each woman played the game three times. The second and third experiments were identical except the store names were changed. The second time, three-digit numbers were used, and the third time three-letter nonsense syllables were used for store names.

WIDTH OF PRODUCT AND PRICE LINES

As stated earlier, the basic research project was oriented toward providing information on those aspects of consumer behavior which can ultimately help in business policy problems. It is convenient to divide marketing policies into four broad areas—product line, advertising, pricing, and selling facilities. The first of these refers to the number and variety of products which is offered for sale by a retailer or produced by a manufacturer. The meaning of the second and third of these categories is self-evident, while the last refers to size of store, number of salesmen per customer, and other related items which go to make up a retailer's store environment.

The present study is primarily concerned with the effects of product line policy on retail sales. The following aspects of this general problem will now be discussed:

1. Number of items to be carried.
2. Depth versus width of assortment.
3. Width or narrowness of price line
4. Number of stores stocking the items carried by a retailer—exclusive dealerships versus the widely handled product.
5. Size of total market for the items stocked.

Product Line Decisions

The number of items handled by a store directly affects that store's attractiveness. On the one hand, the busy, tired shopper does not want to risk entering a store which will very likely not carry the item she wants. For this reason, an increase in the number of items stocked by a retailer will tend to bring in customers. This is the essential attraction of the great department store. On the other hand, an increase in the variety offered by the seller will increase the time and effort needed by the buyer to work her way through many store aisles and counters and find the item she wants. An optimal number of items can only be determined by balancing off these two effects. The elements of an analytic method for solving such a problem have been developed during the course of this basic research program. This operations research technique is, as far as one can determine, the first attempt to deal with the problem systematically, and it is described below in greater detail.²

Two stores carrying the same number of items can differ considerably in the variety of items they offer. One store may offer for sale a wide variety of clothing, and stock a fairly small selection of dresses among its many items. Such a retailing unit is said to carry a *wide* product line. On the other hand, another store may have as many items as the first but by offering for sale nothing but dresses, it can carry a very large selection of these garments. Such a specialty store's product line is characterized by its *depth*. The effect on sales of depth and width of product lines and of the width of choice offered the consumer on price line was investigated experimentally and by interview and the results are presented later in this report.

The same is true of the remaining two subheads. These refer to the choice which dealers must make between selling widely demanded items already offered for sale by many other dealers (as does a liquor seller who stocks mostly popular brands of whiskey and gin) and specialization in items and brands which are less widely sold and less frequently demanded by the body of consumers as a whole (as in the case of a dealer in fine wines who sells primarily to connoisseurs).

Taken in general terms these are the four major decisions on product line which must be made by a dealer—the number of items to be sold, width versus depth of the assortment, the width of price range to be handled, and popularity and competitiveness of the line to be carried. Even when these policies are decided upon, the important questions of detail, the specific items to be carried, remain to be determined. But especially

² The original version included a discussion on optimal variety in retailing. This section consisted of the development of a mathematical model, which while quite relevant to the problem was not germane to the design, implementation, or reporting of the experiment, and was therefore omitted.

in a large enterprise, a department store or a mail-order house, only the broader questions can profitably be answered by top management. The brand and number of handkerchiefs and blouses to be ordered must be decided by individual departments and often by individual buyers. The questions to which this study has been directed, then, represent some of the most important decisions facing top management in retailing organizations.

To examine these questions the shopping games, which have already been described, were undertaken. They were designed to provide an experimental situation which could isolate the effects of a retailer's product line decisions on the demand for his products. We turn now to a description of the results for each of these decision problems in turn.

Width versus Depth of Assortment

To examine the relation of game behavior and the width or depth of the store's product assortment the following procedure was employed. The stores were classed in their three categories of wide assortment, medium assortment, deep assortment. Since the number and size of stores of each variety was the same, the total number of shoppers and purchasers in each type of store could be taken to be indicative of its popularity.

Before giving the results of this tabulation it is necessary to point out that the results are not wholly reliable. For example, while stores carrying a wide product line turned out to be most popular when taken as a class, not all the stores carrying a wide selection were more popular than all stores carrying a narrow selection. This problem arises in all the tabulations contained in this chapter. To a considerable extent this is to be expected because other store characteristics besides width of product line were permitted to vary. If no such phenomenon had been observed it would have implied that width of product line was the only one of these store characteristics having an important influence on sales—a rather implausible state of affairs. In addition it is suspected that this ambiguity in the results when taken store by store is a consequence of the fact that the experimental procedure did not sharply define the characteristics of the stores in the minds of the panel members. This is one respect in which future experimental procedure can be sharpened up.

Table 2 gives game results of depth and width of product line. The table shows the number of women shopping and purchasing in each class of store during the games.

The two results are pretty much in agreement. They suggest that customers are attracted by stores carrying a wide assortment. A very deep assortment is also fairly attractive. But the relatively colorless store which carries a medium line is the least attractive of the three.

These results can be compared with actual shopping behavior of the same panel members as reported in interviews. Their recent purchases in downtown Philadelphia stores were recorded. The stores were then clas-

TABLE 2

NUMBER OF GAMES SHOPPERS, AND PURCHASERS BY DEPTH,
WIDTH OF PRICE LINE

<i>Product Line</i>	<i>Number of Shoppers</i>	<i>Number of Purchasers</i>
Deep.....	228	102
Medium.....	182	81
Wide.....	286	111

sified rather impressionistically into the wide, medium, and deep categories. Naturally the quantities purchased would be affected by the size and number of the stores in each category. For example if there were many more big stores in the medium category most sales would naturally be expected to fall in this group. This figure would not, therefore, represent the effect of width on sales. Moreover a store cannot measure its success in terms of sales alone. Rather its sales must be related to the magnitude of its investment. This procedure was followed here. The figure finally computed was a rough approximation to sales per dollar of investment, rather than unadjusted sales, and the purpose was to examine the effect of width of assortment on the former figure.

It must be re-emphasized that this adjustment was not required for the games since each category in effect included nine stores of identical size. That means that investment in each category was for practical purposes the same, so that the denominator in $\frac{\text{sales}}{\text{investment}}$ did not vary from category to category. That is to say sales and $\frac{\text{sales}}{\text{investment}}$ varied in precisely the same manner. Table 3 summarizes these results.

TABLE 3

ACTUAL $\frac{\text{PURCHASES}}{\text{INVESTMENTS}}$ BY WIDTH OF PRODUCT LINE

<i>Product Line</i>	<i>Index of $\frac{\text{Purchases}}{\text{Investments}}$</i>
Deep.....	0.52
Medium.....	0.42
Wide.....	0.52

Here deep and wide product lines appear to be equally effective sales producers, but once again the undecided medium store suffers relatively.

Width of Price Line

Similar procedures were adopted to examine the effects of width of price line. In the games there were three classes of stores when classified by width of price line. The store with the narrowest price line carried items ranging in price from \$6 to \$8. The medium wide price line store

carried items ranging from \$5 to \$9, while the remaining stores stocked goods whose prices varied between \$3.50 and \$10.50. It will be noted that all three sets of stores offered approximately the same average price level. The price variation that was permitted was only variation in price range. Table 4 contains the shopping game results classified by width of price line.

TABLE 4
NUMBER OF GAMES SHOPPERS, AND PURCHASERS BY WIDTH OF
PRICE LINE

<i>Price Line</i>	<i>Number of Shoppers</i>	<i>Number of Purchasers</i>
Narrow.....	226	91
Medium.....	220	94
Wide.....	260	109

Here a medium and narrow price line seem about equally unattractive. A narrow price line attracts a few more shoppers but somewhat fewer purchasers than does an in-between store. In both cases there is a considerable advantage to the store with the wide price line.

This can again be compared with the results of the interviews on actual shopping behavior of the panel members. Again the results are based on a rough classification of downtown Philadelphia stores, this time by width of price line. These are given in Table 5.

TABLE 5
ACTUAL $\frac{\text{PURCHASES}}{\text{INVESTMENT}}$ BY WIDTH OF PRICE LINE

<i>Price Line</i>	<i>Index of $\frac{\text{Purchases}}{\text{Investments}}$</i>
Narrow.....	0.47
Medium.....	0.43
Wide.....	0.54

The results are again fairly consistent with those of the games. Stores with wide price lines are by far the most attractive to purchasers while this time the medium wide price lines are the poorest sellers.

Exclusiveness of Product Line

There remains the question of whether it pays a store to avoid competition by not selling products which everyone offers because these items are in greatest total demand. Here, unfortunately, the game experiments turned out to be unsatisfactorily designed, inadequate attention having been given to the matter in the planning of the research procedure. It was consequently difficult to obtain a clear-cut classification of the stores in accord with the present question. Nor was it easy for panel members to distinguish among stores for this characteristic. We present such results as were obtained in Table 6.

However it is hardly surprising that there is little agreement of these with the results of the interviews.

TABLE 6

NUMBER OF GAMES SHOPPERS, AND PURCHASERS BY EXCLUSIVENESS OF STORE LINE

<i>Degree of Exclusiveness</i>	<i>Number of Shoppers</i>	<i>Number of Purchasers</i>
Competitive.....	215	89
Medium.....	195	94
Exclusive.....	286	111

This suggests that exclusive stores are likely to be most attractive to customers, with the in-between stores again the least effective sellers.

The interview results on actual shopping are shown in Table 7.

TABLE 7

ACTUAL $\frac{\text{PURCHASES}}{\text{INVESTMENT}}$ BY EXCLUSIVENESS OF STORE

<i>Degree of Exclusiveness</i>	<i>Index of $\frac{\text{Purchases}}{\text{Investment}}$</i>
Competitive.....	0.61
Median.....	0.45
Exclusive.....	0.30

These results are in rather sharp conflict with the game results. They argue against the advisability of exclusiveness although giving a fairly low score to the in-between store.

Conclusion: The In-between Store

There is no need to labor the meaning of the preceding results. Most of the choices in question involve little difference in retailer cost. Where costs are not affected, the retailers best strategy on width of product line, and so on, is to make the choice which maximizes sales per dollar of investment after the sales are weighted by the appropriate profit margins. This chapter has indicated some of the ways in which sales may be affected by the choices in question.

One result, however, does appear to stand out. The in-between store seems most often to come off poorly. Customers would appear to be attracted by stores of distinct character rather than retail outlets which give the appearance of being unable to make up their minds. As is so often the case, clear-cut personality and decisiveness would appear to be good business.

SHOPPING TO GATHER INFORMATION

Consumer Information and Rational Choice

Economic theorists have sometimes disputed the meaning and validity of the assumption of "rational" consumer behavior. In most economic theoretical analysis, it is convenient and useful to assume that consumers know just exactly what their preferences are and go out and spend their

money so as to get the most out of each dollar in terms of those preferences. Now it is clear that this is not an accurate picture of reality. Buyers rarely know exactly what they want and they are often poorly informed about the places, prices, and qualities in which these items can be obtained. Thus the rationality assumption will often be violated in practice. But there is some confusion about the way in which this violation is likely to occur. For the practical marketing man, knowledge of the buyer patterns of irrationality can be of the utmost importance.

Evidence of Low Information Gathering in Shopping Behavior

Recently several studies including that described in this report have brought to light materials which can be interpreted as evidence of consumer irrationality. Noteworthy is the study of the Survey Research Center of the University of Michigan³ whose interviews indicate that the consumers rarely go through protracted deliberative procedures before making an expensive purchase like a TV set, a refrigerator, a washing machine, or a stove. Nearly half the people interviewed visited only one store before making their purchases. Less than a quarter of the buyers remembered receiving information from advertisements. Frequently little or no family discussion was reported to have occurred before the purchase. One third of the buyers consulted no more than one source of information.

Alderson and Sessions' basic research program has produced results which conform closely with those just cited. On purchases of less expensive items like underwear, hats, cooking utensils and toys, purchasers almost never investigated more than one retailer. Our consumer panel of 61 housewives were interviewed every other week for a 10-week period and asked to describe in considerable detail their shopping trips. The interview covered the prepurchase considerations, the actual course of the shopping trip, and the degree of post-purchase satisfaction with the items purchased. Naturally, many of the goods purchased consisted of fairly routine household purchases, any one of which constitutes a very small proportion of the families' expenditures. But since so many of these items are bought in most families, the total cost of these purchases may loom large in the family's budget. The interviews indicated that in the vast majority of cases, the housewife made her purchase in the first store she entered. The data are presented in Table 8.

While, as is to be expected, there is somewhat more shopping around prior to the completion of a major purchase, even these expensive items are bought with little store-to-store investigation. Thus, even in purchases of refrigerators, television sets, automobiles, and furniture half of the purchases were made after visiting one store.

Our interviews indicated somewhat more prepurchase discussion than

³ George Katona and Eva Mueller, *The Dynamics of Consumer Reaction* (New York: N.Y.U. Press).

TABLE 8
NUMBER OF STORES ENTERED IN MAKING PURCHASES

<i>Number of Stores Entered</i>	<i>Percent of Total Purchases</i>	
	<i>Minor</i>	<i>Major</i>
Percent of purchases.....	100.0	100.0
One store.....	87.4	50.0
Two stores.....	6.1	14.1
Three stores.....	1.1	9.0
Four or more stores.....	5.4	26.9
Total number of purchases in the sample.....	627	78

was reported by the University of Michigan Survey Research Center. But they suggested that much of this discussion was sporadic and not carefully directed at information gathering. Price was certainly not a leading matter in these conversations. Frequently the prepurchase discussion was simply used as an outlet for griping—as an opportunity to express dissatisfaction with current possessions.

It also appears that the care and deliberation with which consumers spend their money increases with size of income; for example, those who can least afford to waste money seem to take the least trouble gathering information.

Alternative Sources of Information⁴

There is some danger that these results may be misleading. Careful consideration shows that while they do suggest rather extreme irrationality in shopping behavior, this is not necessarily so. Rationality, that is efficiency in instrumental behavior, requires that the costs as well as the benefits of any alternative be taken into account. Information is clearly useful in making shopping decisions. But the acquisition of that information can be exceedingly costly in time and effort which can ill be spared by an exhausted housewife.

Especially in the case of a minor purchase, the few cents that a shopper can save by shopping around may simply not be worth the time and effort required on any reasonable weighing of the alternatives. It is true that minor purchases are frequent and in the long run small savings on each purchase can add up to a formidable total. But by the same token in such purchases the housewife may be a more efficient shopper if she spreads her information gathering over time. By buying for a while in a few different stores she may soon learn their characteristics. As a result she will be able to purchase in the first store she enters confident that this is the retailer who is very likely to suit her needs best.

The One-Store Shopper

One important practical conclusion that can be drawn from the foregoing discussion is that the shopper who enters a store is also likely to purchase there. This means that there is considerable justification for retailer strategy which is designed to entice customers into the store. Loss

leaders, entertainment for the children, and other such devices can be expected to increase profitable sales and not just attract visitors.

But by and large a store attracts customers by its more durable and less superficial characteristics. Our study indicates that a department store shopper has a tendency to rely on a single store and to shop at other stores as an occasional check on her basic choice or in response to bargain offers. Her preferred store is one which reflects a view of contemporary life consistent with her own; one which offers the merchandise that seems to belong together and fits in with her home or her vision of what she wants her home to be. A conviction that competing department stores have definite personalities has been registered strongly in our intensive study of shopping behavior.

Store Clusters

Closely related to this phenomenon of shopping by store personalities is a manifestation which we call shopping by store clusters. Our research indicates that customers prefer to do their shopping in stores which are geographically proximate and similar in character.

For example, a group of the women interviewed in connection with the basic research program was asked to indicate first and second downtown Philadelphia department store preferences. The stores were then grouped into two major pairs. One which we will call Cluster A included a set of stores which were close geographically and fairly homogeneous in character. The other group, Cluster B, consisted of the only other major set of geographically proximate stores—but this is a set of stores differing widely in character. The results are given in Table 9.

TABLE 9
FIRST AND SECOND STORE CHOICES BY GEOGRAPHIC STORE CLUSTERS

<i>If First Choice Store in Cluster A</i>			<i>If First Choice Store in Cluster B</i>			<i>Total Number of Cases</i>
<i>Second Choice Store Also in A</i>	<i>Second Choice Store Not in A</i>		<i>Second Choice Store Also in B</i>	<i>Second Choice Store Not in B</i>		
<i>Total</i>			<i>Total</i>			
26	21	5	13	0	13	39

These results suggest that a shopper's second choice store tends to be in the same neighborhood as the first choice store if there are nearby stores similar in character to the first choice store. Geographic proximity alone is not enough to hold shoppers. Of the 26 women whose first choice store was in Cluster A, 21 made a second choice in the same cluster,

while of the 13 women whose first choice store was in Cluster B all made a second choice outside that cluster.

But similarity in store personality without closeness will also not succeed in keeping customers within a store cluster. The stores were reclassified into Cluster C (very much like A) and another cluster, D, consisting of stores similar in character but not very close together. The results analogous with those of Table 9 are shown below in Table 10.

TABLE 10
FIRST AND SECOND STORE CHOICES BY STORE PERSONALITY CLUSTERS*

Total	If First-Choice Store in Cluster C		Total	If First-Choice Store in Cluster D		Total Number of Cases
	Second Choice Store	Second Choice Store		Second Choice Store	Second Choice Store	
	Also in C	Not in C		Also in D	Not in D	
27	20	7	14	1	13	41

* The total in Table 3 is somewhat greater than that in Table 2 because several downtown stores which readily fitted in one of the personality clusters could not conveniently be classified in a geographic cluster.

These results are really quite in accord with what may be expected in advance. But it does suggest a somewhat more startling conclusion; that in department stores close competition may help rather than hinder sales.

FAMILY INTERACTION PATTERNS

Qualitative Nature of the Material

It is not easy to quantify the processes whereby purchase decisions are affected through the interaction of family opinions and desires. Statistics on the number of times a purchase plan was discussed between husband and wife and other similar numerical data are not likely to be very informative. They cannot indicate the intensity of the discussion, the range of topics touched upon, or the degree of influence on the final decision. Some of the collected statistics that have bearing on this point are presented below.

But it is also appropriate to include a brief summary of the qualitative impressions obtained from depth interview data, group discussion with the interviewees, their responses to Thematic Apperception Tests, and other subject material.

The Alderson and Sessions shopping panel of 61 housewives which was subjected to intensive testing, interviewing, and clinical game experi-

mentation was chosen so as to maximize the opportunity to observe interaction processes. Women were chosen in pairs; the first women selected were each asked to name a close friend with whom she shops at least occasionally. This friend was then also invited to become a member of the panel. In every case this invitation was accepted. In addition there was one session which husbands were asked to attend and to join in the discussion. A high proportion of the husbands did come and did participate.

Identification of Decision Makers: Minor Purchases

The role of the husband and wife in purchasing decisions varies considerably between minor, everyday purchases and major purchase decisions. In the buying of food and clothing for most family members the housewife acts as the family purchasing agent and the decisions are hers alone. Of course her decisions are influenced by what she knows of family likes and dislikes and it is difficult to tell how her decision pattern on minor purchases changes as a result of family discussions. An extreme case would be that where the husband and children simply refuse to eat some food which she has prepared. Even though the final purchase decision appears to be hers alone, family interaction will have played a major role in persuading her to discontinue her purchase of this item. Thus the housewife, though she ordinarily determines which minor items will be purchased and where they will be bought, is influenced by her family in two different ways. First, she normally wishes to buy things which will please other members of the family insofar as this does not conflict with her own ideas of what is right, proper, healthy, and aesthetic. Second, the other family members can, through their behavior, exert pressure on her to cajole her into buying in accord with their desires. Their importance is not reduced by the fact that both these influences are often exercised subconsciously and manifested in disguised forms.

The predominance of the housewife as the purchasing agent, though not necessarily as a decision maker is indicated in Table 11. Aside from the purchase of major items, which is to be discussed presently, the only

TABLE 11
FAMILY MEMBER WHO SHOPS FOR SELECTED FAMILY PRODUCTS
(Number)

<i>Product</i>	<i>Total</i>	<i>Wife Shops</i>	<i>Husband Shops</i>	<i>Both Shop</i>	<i>Wife Shops— Husband Goes on Final Purchase Trip</i>	<i>Undetermined</i>
Household.....	52	31	0	19	2	0
Food.....	52	36	5	11	0	0
Wife's clothes.....	52	47	0	4	1	0
Husband's clothes.....	52	10	22	17	1	2
Children's things.....	52	40	0	7	0	5
Major items.....	52	8	2	32	9	1

clear exception is the purchase of the husband's clothes in which he himself plays an important role. In nearly half the cases listed he shops entirely by himself. Nevertheless, it is noteworthy that the wife is along presumably in at least the role of consultant and critic in a majority of the cases, and in 20 percent of the families questioned the wife does even this part of the shopping by herself.

The number of other items in whose purchase the husband participates may somewhat overstate the facts. It may be surmised that in a number of these families the husband comes along on a shopping trip for food and household items to help carry things home rather than to participate in the actual purchasing process.

INTERACTION IN MAJOR PURCHASING DECISIONS

The situation changes radically when the family turns from its everyday purchases to the procurement of major items like automobiles and expensive kitchen appliances. This has already been indicated in Table 10. There it is shown that the wife rarely shops for major items by herself. In over 80 percent of the families interviewed these goods are shopped for or bought jointly. However, there is particular reason for treating these figures with caution. Our depth interviews suggested that answers on major purchase decisions tended to be influenced by the interviewees' ideas as to what was proper and conventional. For example the husband was often given credit for judgment on mechanical details and operation of equipment but further probing indicated that this sometimes had little basis in fact and was asserted largely because it represented the respondent's idea of the proper masculine role. Nevertheless husbands do appear to play an important part in major purchase decisions. It is noteworthy that this appears to be true even of purchases of kitchen appliances.

SUMMARY

The major interaction, as is to be expected, is between husband and wife. In some families, however, the decisions are completely in one hand or the other. Where the wife does all the shopping, she is likely to try to do a better job on major items in an effort to please her husband and to assure herself that she is a good shopper. The decision process for appliances usually starts with the wife, but her husband is given credit for examination of the mechanical features and price, and the woman decides on style and performance characteristics. There is little recall or conscious knowledge of effects by neighbors, family, and friends, but there is indirect evidence that these play a role—sometimes significant—in the decision to buy a particular brand.

The importance of this for the practical marketing man is clear. For it indicates to whom he must direct his appeals in getting people to purchase his products. It suggests for example that advertising of less expensive items which are not specifically designed for the use of men, that is, items other than men's clothes, shaving apparatus, and so on, can afford to emphasize feminine appeals considerably more than advertising of major purchase commodities.

A Tasting Experiment*

.....

NORMAN THEODORE GRIDGEMAN†

Introduction

SUCROSE IS SWEETER THAN GLUCOSE BY A FACTOR THAT HAS A PERSONAL equation and whose usual value has been variously estimated to fall in the range of about 1.2–2. Glucose, in the commercial form known as corn syrup or confectioner's glucose, possesses certain desirable qualities as an ingredient of jam. Apart from the matter of cost, there is then an open question as to whether it is a good thing to use some glucose in jam manufacture; in other words, how would the customer—and in particular a customer accustomed to all-sucrose jams—react to it? This question was recently posed in Ottawa, with special reference to a sugar mix of three parts sucrose and one part glucose, which mixture makes a jam that is certainly not obviously different in taste from a standard jam. It was decided to assay the difference on a laboratory panel consisting of a few dozen local volunteers. The design and analysis of this trial are the subject of the present paper.

The original plan called for four independent manufacturers (hereafter referred to by the first four Roman numerals), each to make parallel batches of jam from the same raw materials. One, the “regular” batch, was to contain sucrose only, and the other, the “glucose” batch, was to differ from the first by a weight/weight replacement of twenty-five per cent of the sucrose by glucose. This was to be done with two kinds of jam: strawberry and raspberry. Thus $4 \times 2 \times 2$ jam samples were expected. In the event one manufacturer did not prepare any raspberry jams, so there were fourteen samples to deal with.

Information was sought on two aspects of the regular: glucose comparison, namely sweetness and preference. Forty-eight unselected subjects (thirty-nine men; nine women) from the divisional staff were recruited as tasters. Each was presented with small samples (about five grams in a glass dish) of the jams in the form of seven coded pairs and

* Reprinted from *Applied Statistics*, Vol. V, No. 2 (June, 1956), pp. 106–12. (NRC No. 3850.)

† Division of Applied Biology, National Research Laboratories, Ottawa.

was asked to indicate on a check sheet which member of each pair he thought was the sweeter and which he preferred. All the pairs were presented together, but the coding key was varied in such a manner as to exclude any overall bias in the order of presentation of the seven jam-pairs or in the left: right disposition of regular: glucose within pairs. Within that restriction the order was randomised. The tasting was done in air-conditioned single-seater booths under neutral light. Unsalted crackers and drinking (or rinsing) water were available.

Results

A summary is given in Table 1 of the way in which the forty-eight persons voted with regard to sweetness and to preference. A few "neutral" answers turned up, i.e. no vote for either alternative. Actually, in briefing the subjects we urged them to make a decision no matter how little confidence they might feel in so doing, but some defaulted nevertheless. Only one subject recorded neutrals for all seven pairs. Overall, the glucose jams received slightly more than 40 per cent of the decisions on both sweetness and preference, whereas the regular jams received over 50 per cent. In both instances the vote difference is large enough (the probabilities are beyond the 5 per cent point) to suggest that small but real distinctions exist between the two kinds of jam for these subjects. These are "two-tail" probabilities. However, taking account of our knowledge that the regular jams cannot be *less* sweet than the glucose jams, we may subject the sweetness difference to a "single-tail" test of significance, which will make for a sharper distinction. In fact the probability of a fortuitous difference equal to or greater than that observed (in favour of the regular jams being sweeter) is less than 1 per cent.

Sweetness vis-à-vis Preference

There is no evidence of any general association between sweetness and preference. Take, for instance, the overall proportion of preference for glucose jams: it was 42 per cent among the votes cast for the regular jam as the sweeter and 43 per cent among those cast for the glucose jam as the sweeter. Inspection of the results of the seven individual pairs reveals some variation in this respect, but all of it is within the normal range of random fluctuation. This finding applies to the group, of course, and not necessarily to the component individuals, some of whom in fact do associate preference with greater or less sweetness. In this connection it is of interest to segregate the subjects according to the number of occasions (maximum 7) they expressed a preference for the sample that they also judged to be the sweeter. For this purpose we shall exclude the nine subjects who gave one or more neutral answers. In Table 2 the observed frequency distribution is compared with the binomial expectations on the hypothesis that the association of preference with sweetness is a random variable with $p = 145/273 = 0.531$.

$$\begin{array}{r} 27 \\ \times 7 \\ \hline 273 \end{array}$$

TABLE 1

DISTRIBUTION OF VOTES FOR SWEETNESS AND FOR PREFERENCE AMONG THE 48 TASTERS

STRAWBERRY I					
Prf	Swt				
	R	G	N	Σ	
R.....	21	10	0	31	
G.....	7	6	0	13	
N.....	2	0	2	4	
Σ.....	30	16	2	48	

RASPBERRY I					
Prf	Swt				
	R	G	N	Σ	
R.....	13	18	0	31	
G.....	9	6	0	15	
N.....	0	1	1	2	
Σ.....	22	25	1	48	

STRAWBERRY II					
Prf	Swt				
	R	G	N	Σ	
R.....	18	7	0	25	
G.....	14	7	0	21	
N.....	1	0	1	2	
Σ.....	33	14	1	48	

Manufacturer II did not supply
raspberry jams.

STRAWBERRY III					
Prf	Swt				
	R	G	N	Σ	
R.....	9	11	0	20	
G.....	16	9	1	26	
N.....	0	1	1	2	
Σ.....	25	21	2	48	

RASPBERRY III					
Prf	Swt				
	R	G	N	Σ	
R.....	11	9	0	20	
G.....	13	11	0	24	
N.....	1	1	2	4	
Σ.....	25	21	2	48	

STRAWBERRY IV					
Prf	Swt				
	R	G	N	Σ	
R.....	15	9	0	24	
G.....	10	12	0	22	
N.....	1	0	1	2	
Σ.....	26	21	1	48	

RASPBERRY IV					
Prf	Swt				
	R	G	N	Σ	
R.....	15	11	0	26	
G.....	10	9	0	19	
N.....	2	0	1	3	
Σ.....	27	20	1	48	

GRAND TOTAL					
Prf	Swt				
	R	G	N	Σ	
R.....	102	75	0	177	
G.....	79	60	1	140	
N.....	7	3	9	19	
Σ.....	188	138	10	336	

Swt = Sweetness, Prf = Preference, R = Regular, G = Glucose, N = Neutral. Roman numerals indicate manufacturer.

For example, with strawberry I, 10 persons voted the glucose jam as sweeter but preferred the regular jam.

TABLE 2

SWEETNESS-PREFERENCE ASSOCIATION AMONG THE TASTERS
WHO RECORDED NO "NEUTRAL" JUDGMENTS

<i>Number of Associations in Pair Contrasts (Maximum = 7)</i>	<i>Observed Frequency</i>	<i>Expected Binomial Frequency</i>
0.....	4	0.20
1.....	1	• 1.54
2.....	6	5.24
3.....	8	9.89
4.....	4	11.20
5.....	7	7.61
6.....	6	2.87
7.....	3	0.46
Total.....	39	39

A test for goodness of fit yields $X^2 = 21.0$ for 4 degrees of freedom (after pooling of the terminal pairs of groups), which is well beyond the customary significance levels. (Throughout this paper X^2 will be used to distinguish the calculated statistics from tabular χ^2 .) We therefore infer that some subjects specifically based their preferences on their judgements of sweetness. The overall mean of $145/39 = 3.718$ is close to 3.5, showing that the individual associations cancelled out for the panel as an entity, i.e. there were evidently about as many subjects who preferred the sweeter as preferred the less sweet.

Personal Differences

With regard to the sweetness judgements considered by themselves, there is no evidence of heterogeneity among subjects. The same can be said of the preference judgements. It is, however, more than likely that this is a consequence of the comparatively small size of the trial and small differences between the contrasted jams. A larger trial with greater replication might be expected to reveal differences at present masked. As it is, the distributions of votes-out-of-7 are almost exactly what would be expected if all the tasting had been done by one subject instead of being spread over many and if the seven jam pairs had a common sweetness difference. This can be appreciated by mere inspection of Table 3, in which the observed frequency distributions (excluding subjects who gave neutral answers) are compared with the binomial expectations for a perfectly homogeneous panel whose probability of casting a decision for a glucose jam is $132/308 = 0.429$ on sweetness and $121/280 = 0.432$ on preference. Goodness-of-fit tests support the intuitive conclusion that there is no evidence of heterogeneity: when the 0, 1 and the 5, 6, 7 vote-groups are pooled, giving 3 degrees of freedom, we find X^2 to be 1.95 for sweetness and 0.23 for preference, both obviously below significance.

Differences between Manufacturers and between Fruits

The average distribution of sweetness votes between "regular," "glucose," and "neutral" for the seven pairs (see Table 1) is 188/7, 138/7,

TABLE 3
DISTRIBUTION OF SWEETNESS AND PREFERENCE VOTES, EXCLUDING THE
TASTERS WHO RECORDED "NEUTRAL" JUDGMENTS

Number of Votes-in-7 for Glucose Jam	Sweetness Frequency		Preference Frequency	
	Observed Frequency	Expected Frequency	Observed Frequency	Expected Frequency
0.....	1	0.87	1	0.76
1.....	6	4.58	4	4.06
2.....	7	10.32	9	9.27
3.....	14	12.92	12	11.75
4.....	11	9.71	8	8.94
5.....	4	4.38	5	4.08
6.....	1	1.10	1	1.03
7.....	0	0.12	0	0.11
Total.....	44	44	40	40

and 10/7. If all seven pairs were indistinguishable as regards their sweetness differences, and if all subjects were of equal sensory acuity, we should expect the individual distributions (e.g. 30, 16, and 2 for strawberry I) to conform to the average; the degree of conformity can be judged by χ^2 with $(7 - 1) \times (3 - 1)$ degrees of freedom. Making the appropriate test we find $X^2 = 13.4$, which falls on the 0.33 point of χ^2 with 12 degrees of freedom, and again is not significant. As might be expected from the lack of evidence of differences between all seven pairs, neither is there any evidence of differences between the two types of jam, strawberry and raspberry. The same can almost be said about differences between manufacturers except that, with regard to the preferences, there is a suggestion of such a difference. For example, manufacturer I collected 62 preference votes for his "regular" jam, whereas manufacturer III collected only 40. A χ^2 -test for heterogeneity between the four manufacturers yields $X^2 = 12.3$, which nearly achieves the $P = 0.05$ point of χ^2 with 6 degrees of freedom, namely 12.6.

Differences in General: Another Approach

If we ignore the returns of those subjects who recorded one or more "neutral" votes, we can test for heterogeneity among the remainder by means of Cochran's criterion¹ Q , whose limiting distribution is that of χ^2 . This is a more searching criterion than those used in the preceding sections, although here partially offset by our having to jettison some of the information. The criterion is in principle applicable to any body of data that can be set out in a complete table of M rows times N columns, each entry being 0 or 1, expressive of some simple alternative. Let T be any row total, let U be any column total, and let $\Sigma T = \Sigma U = S$, the grand total. Then

¹ W. G. Cochran, "The Comparison of Percentages in Matched Samples," *Biometrika*, Vol. XXXVII (1950), p. 256.

$$Q = \frac{(N-1)[N(\sum U^2) - S^2]}{NS - \sum T^2}$$

assessed against χ^2 , with $N-1$ degrees of freedom, is a criterion of the heterogeneity among columns of the within-column distributions of 1's. The same expression, with N replaced by M , and with U and T transposed, yields another Q that, assessed against χ^2 , with $M-1$ degrees of freedom, is a similar criterion of row differences. Applying these methods to the data at hand we obtain the results given in Table 4, in which

TABLE 4

PROBABILITIES OF HOMOGENEITY: (i) AMONG THE
SENSORY DISCRIMINATION OF THE TASTERS; (ii)
AMONG THE DIFFERENCES IN THE PAIRS OF JAMS

	<i>Sweetness</i>	<i>Preference</i>
Number of tasters	44	40
Number of jam-pairs.....	7	7
Among tasters	$\{Q \dots 44.4$	42.4
	$\{P \dots 0.43$	0.34
Among jam-pairs	$\{Q \dots 7.5$	14.3
	$\{P \dots 0.28$	0.027

Q is Cochran's index, whose distribution is approximately that of χ^2 .

it may be seen that only one Q , that of preference differences among the seven jam pairs, is suspiciously high. This links with the suggestion emerging in the previous section that there may be some difference between manufacturers as far as these batches of jam are concerned. Whether this would be borne out by new batches is problematic.

Discussion

In interpreting the results we must take into account the fact that they stem from a laboratory sensory-difference trial, involving 336 paired comparisons. Even under conditions of such stringency only a small difference emerged. It is accordingly unlikely that the public at large, in normal (non-comparative) jam-eating practice, would sense any difference between regular and glucose jams in so far as the samples used in the present investigation are typical samples.

An interesting point emerging from the trial is the absence, overall, of any correlation between sweetness and preference, and the fact that this is mainly due to a cancelling out of individual correlations. For some subjects did associate sweetness and preference, but there were about as many who favoured the sweeter as favoured the less sweet (according to their own taste decisions). This finding is, of course, conditioned by the particular magnitude of the sweetness difference. Clearly, if one jam was of normal sweetness and the contrasted jam was unsweetened, few would prefer the latter. Moreover, the cancelling out referred to is applicable only to the population represented by this panel of adults, drawn

from a group of youngish scientific workers, and it would be presumptuous to generalise further.

Acknowledgments

The author would like to thank the Department of Agriculture, and the Food and Drug Directorate of the Department of National Health and Welfare, for permission to use the experimental data.

An Experimental Method for Estimating Demand*

.....

EDGAR A. PESSEMIER†

A GENEROUS INCREASE IN THE AVAILABLE EMPIRICAL DATA CONCERNING demand for individual branded products would be of considerable value to economists and businessmen. For the economist these data would provide an important foundation of fact upon which the structure of microeconomic price theory could rest. For the businessman they would yield helpful indications about how buyers evaluated his brand, as well as the brands of competitors, thereby removing some of the guesswork from decisions concerning price, product design, and promotional activities. Consequently, it is interesting to find that relatively little has been done to obtain demand schedules for individual branded goods.¹ Why has this been the case? The answer can be found principally in the problem of measurement. When an attempt is made to estimate demand under market conditions, an extended period of observation is required, and the cost of gathering data is often high. Furthermore, the use of a protracted period of observation introduces a variety of uncontrolled variables whose effect cannot be accurately isolated and assessed. It appears that, so long as the market is used as the source of data, there is little hope of overcoming these difficulties.

The Experimental Approach

Hope is held, however, for obtaining simple approximations of the demand for branded goods by gathering information about the behavior

* Reprinted from the *Journal of Business*, Vol. XXXIII, No. 4 (October, 1960), pp. 373-83 by permission of the University of Chicago Press. Copyright 1960 by the University of Chicago, all rights reserved.

† State College of Washington.

¹ For a summary of much of the work done in this area, see Edward R. Hawkins, "Methods of Estimating Demand," *Journal of Marketing*, April, 1957, pp. 428-38.

of buyers in a controlled environment.² If the length of time between buying decisions is greatly reduced, buyers' actions can be observed without having to evaluate such disturbing factors as changes in the branded product, its promotion, its method of distribution, and its competition, or changes in the economic or psychological characteristics of the industry's buyers. By this procedure *all important influences on the buyer, except price, can be held constant* so that the independent effect of changes in price can be observed. The crucial problem that must be dealt with when research is conducted in this manner is the preservation of a sufficiently realistic situation to insure that subjects will respond in the experimental setting in approximately the same way they would in the marketplace: the state of *ceteris paribus* must include as a necessary condition an experimental environment that is not unworkably artificial.

Since this discussion is limited to an analysis of the demand for consumer goods of relatively modest unit price, the personal experience of the buyer of such goods is easy to describe. He seeks to satisfy his wants by purchasing goods from existing institutions and assortments, and he has limited time, information, and funds to use in gaining these ends. By the acts of gathering satisfactions in the market, he is expressing personal judgments about the relative value of what the market has to offer. When taken over a given period of time, the sum of the preference-motivated actions of all buyers represents demand. In other words, within the limits of the consumer's capacity to act, demand for a product depends on how consumers evaluate the product's relative worth. Since in the market it is often difficult to determine the demands or preferences for branded products over a moderate range of price variation, the question naturally arises: Can it be done in a controlled environment? An affirmative answer can be given *provided* the buyer can be placed in a position where the consequences of his actions in the experimental environment will have an impact on his well-being and conduct similar to what they would have in the market: the experimental conditions should be *psychologically equivalent* to the market, not necessarily physically identical. If the experimental situation is made "real" by duplicating those aspects of the market which influence buyer action, then the experimental results will closely parallel the decisions made by consumers confronted by similar conditions in everyday life.

The experiments reported here were designed to accomplish this end by having subjects go on simulated shopping trips. As it would have been on a real shopping trip, the subject's goal was to maximize the satisfaction

² During 1957 three reports appeared concerning important groups of game experiments: Cyeil C. Herrmann and John B. Stewart, "The Experimental Game," *Journal of Marketing*, July, 1957, pp. 12-30; *Basic Research Report on Consumer Behavior* (Philadelphia: Alderson & Sessions, April, 1957), pp. 1-01-6-05; and Donald Davidson and Patrick Suppes, with Sidney Siegal, *Decision Making* (Stanford, Calif.: Stanford University Press, 1957); see also Edgar A. Pessemier, "A New Way To Determine Buying Decisions," *Journal of Marketing*, October, 1959, pp. 41-46.

he could obtain from prevailing market conditions and limited funds. Each participant in the experiment was told how much money he had to spend, the assortment of brands available in each class of goods from which he was to make a purchase, and the price of each item. In every case the subject made purchase decisions from assortments that contained goods which he purchased frequently for his own personal use. The participant in the experiment could maximize the satisfaction he might obtain on any one of the simulated shopping trips by selecting the brands he preferred in light of the price at which they could be purchased. Since he had a stated sum available, the act of making the selections also determined the amount of change he would receive. The experiment was administered to groups ranging in size from twenty to fifty subjects, and at the conclusion of the experiment one member of each group, selected at random, actually received the merchandise and change called for by his decisions during one of his simulated shopping trips. It seems fair to state that a reasonably close parallel to real shopping conditions was maintained during the experiment and that useful information was obtained about consumer behavior.

Although maintaining psychological equivalence was one of the principal objectives of the experiment, it should be noted that it is not essential that subjects respond precisely as they would in a real market environment. It is necessary only that any deviation which may exist be predictable. Had the study reported here been designed to do more than explore the potential value of the experimental method in deriving schedules of demand, deviations in behavior could have been examined under a number of experimental procedures. As additional experiments are undertaken, the method employed should be varied to gain sharper insights into the impact on experimental subjects of such factors as the form of presentation, the procedure followed in modifying price, the types of incentives offered, the number and complexity of the purchase decisions, and the time allowed to complete a simulated shopping trip. If a larger sample representative of the population of buyers of each class of goods is used in a future study, an examination of the behavior of various classes of buyers would also be a promising area for exploration. For example, it could be instructive to examine demand schedules for groups of subjects possessing distinct personal, social, and economic characteristics.

Design and Administration of the Experiment

The experiment was administered to 228 students at Washington State University during the spring of 1959. Although convenience in handling groups was an important consideration in selecting the experimental subjects, it was possible to obtain subjects representing all social class levels and a wide range of fields of major interest. However, a higher proportion of males, upperclassmen, and business administration majors

were present than would be expected in a random sample drawn from the population of approximately six thousand resident students. An extensive statistical analysis of the effect of characteristics of buyers was beyond the scope of this study, but a limited check was made which failed to uncover significant bias introduced by the particular composition of the experimental subjects.

Before the experiment began, the subjects were polled to determine whether they purchased items for their own use from four classes of goods—toothpaste, cigarettes, toilet soap, and headache remedies—and, if they did, what brands they customarily purchased. In addition to the form used to gather these facts, two sets of assortment sheets, or lists of brands, were compiled and duplicated in advance. These sheets listed seven brands of toothpaste, ten brands of cigarettes, eleven brands of toilet soap, and six brands of headache remedies and included all the brands available at the student bookstore in these classifications.³ Because of the convenient location of the student bookstore, the particular lines of merchandise that it stocks, and the absence of effective competition, student patronage was very high. Since each assortment that was used paralleled one found in a retail store in which the subjects frequently shopped, presumably subjects were familiar with the brands and their usual prices.

On the basis of the brand preferences indicated by the subjects, it was possible to modify the prices of each subject's preferred brands on sets of assortment sheets.⁴ By raising the price of the subject's preferred brands on each of a number of assortment sheets, subjects were offered their preferred brands at various increases in price while the other brands remained available at the regular price. By entering the regular price of a subject's preferred brands on a series of assortment sheets on which the prices of all brands had been reduced, it was possible to offer *all* but the subject's preferred brands at various reductions in price. Other than these changes in price, the conditions under which the subjects made their buying decisions were unaltered. As a result, subjects were given assortments from which to make selections on simulated shopping trips that *required the subject to decide whether he would buy the brand he preferred or whether he would change his usual purchasing behavior because of the difference in price.*

When the two series, containing a total of twenty simulated shopping trips, had been prepared for each individual, the experiment was administered to groups of subjects. The subjects were shown samples of the merchandise contained in the assortments, given an appropriate series

³ Each classification also carried one or two additional items labeled "a new brand"; this was included in the assortment to measure the tendency of consumers to buy a new brand when switching brands.

⁴ To eliminate positional bias, the positions of brands in a column and the positions of sheets in a series were randomized.

of assortment sheets, and told that they had \$1.75 available on each shopping trip. This sum was large enough to permit a subject to purchase the highest-priced item in each of the four classifications, if he chose to do so, and still receive change. Although the subjects were requested to assume that they had a current need for the merchandise included in the assortments, they were permitted to postpone making a purchase if they would walk a block to investigate the offerings in another store. As a result, subjects were expected to purchase a fixed number of items, but they were given an opportunity to shop.⁵ In addition, the subjects were told that at the conclusion of the experiment one member of the group, chosen at random, would receive the actual merchandise and change called for by his decisions on one of his shopping trips. Finally, the subjects were asked to make the twenty simulated shopping trips, selecting those items on each trip which would give them the greatest satisfaction from a mix of merchandise and change.

Several additional precautions were taken to reduce any tendency subjects might have to try to win approval by acting "rationally." First, subjects were given no indication of what "rational" conduct might be. Second, they were asked to shop at the same speed and react in the same manner as they would on a real shopping trip. And, finally, subjects were allowed to handle the assortment sheets on which they recorded their buying decisions in a manner which would prevent the identification of a set of responses with a particular individual.

On the average, the experiment was explained and administered to a group in less than thirty minutes. If it had been practical to assemble all subjects in a single location, the nearly fifteen thousand buying decisions which were recorded could have been recorded in little more than a half hour.⁶ The important point is that a large amount of data about the behavior of consumers was accumulated rapidly at low cost. Even if subjects had been handled singly, at least one hundred buying decisions could have been recorded in an experimental environment in less than one-half hour.

In the course of the experiment it developed that five individuals were members of more than one group. As a check on consistency, they were permitted to take part in two sets of simulated shopping trips, several hours to several days apart. About 85 per cent of the decisions recorded on the second series were in direct agreement with the decisions made during the original series. And, in almost all cases where disagreement existed, the magnitudes of the differences were small.

⁵ Because of the type of goods purchased, it was not surprising to find that only three subjects elected to investigate a second set of assortments.

⁶ All 288 subjects made 20 simulated shopping trips and could have bought one item from each of four classifications on each trip. Some subjects, however, did not buy goods from all four classifications, since they did not normally either buy or use the goods included in the classifications, that is, cigarettes for the nonsmoker.

Special Use of "Demand"

As a consequence of the methods employed in the experiments, the terms "demand curves" and "demand schedules," although similar to those used in economics, are used here in a special sense. For example, with reference to Brand A toothpaste, Figure 1 shows that 81 individuals bought Brand A at its regular price of 31 cents. When the price of Brand A increased to 32 cents and all other brands remained at their regular price, 31 cents, 67 subjects continued to buy Brand A, and 14

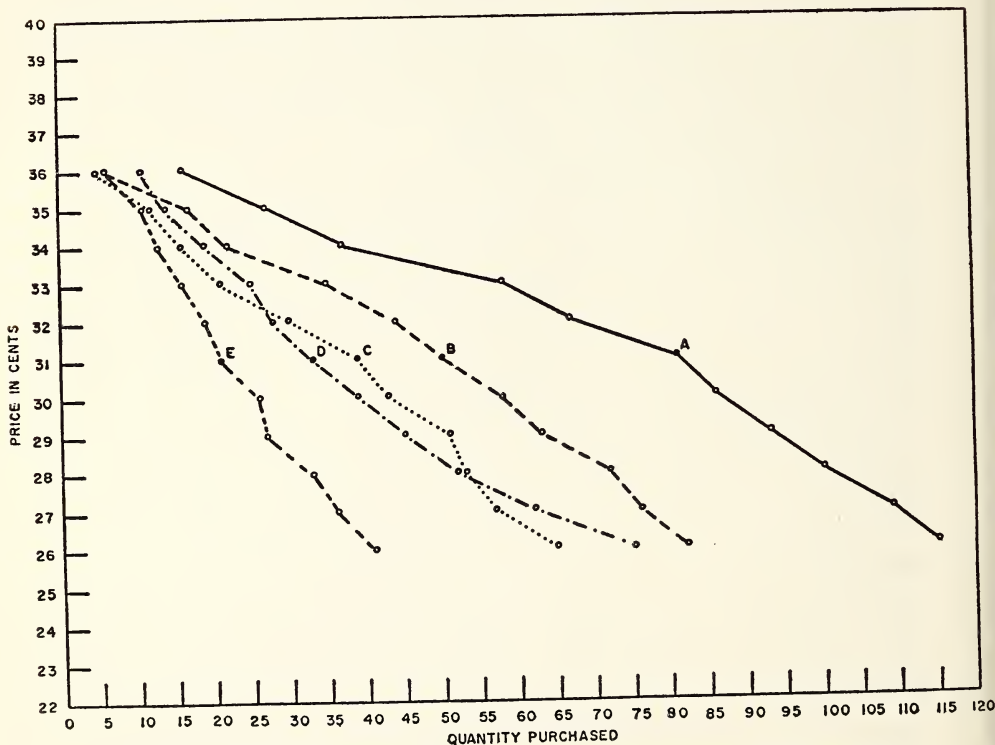


FIGURE 1. Demand curves for toothpaste brands preferred by 20 or more subjects.

switched to other brands of toothpaste. A like interpretation may be given to the remaining points on the upper portion of Brand A's "demand curve" as well as to the upper halves (above solid dots) of all the "demand curves" in this article.

The bottom half of the "demand curve" for Brand A was not obtained by simply lowering the price of Brand A. To do so would have required a four- to fivefold increase in the number of buying decisions made by every subject, since it would have been necessary to reduce individually the price of six brands of toothpaste, ten brands of toilet soap, and so on, through a full range of price reductions. Cutting the number of classifi-

cations included in an exploratory study or risking excessive fatigue on the part of subjects was undesirable. Therefore, the alternative of simultaneously reducing the price of all but the subject's preferred brand was adopted. The effect of this procedure was to eliminate the advantage a brand would have by being the only one available at a reduced price. In the case of Brand A, Figure 1 shows that, among those subjects who normally purchased brands other than Brand A, there were five individuals who would switch to Brand A if it and all other nonpreferred brands were reduced to 30 cents. If Brand A had been the only brand offered at this one-cent reduction in price, additional switching to Brand A might have occurred. Under these circumstances switching would not have been the result of a secondary preference for Brand A so much as the result of the more direct price appeal possessed by Brand A when compared to other brands. In other words, the bottom halves of the "demand curves" understate the effect of price reductions, and the understatement should be more pronounced as the magnitude of the reduction increases.

For some purposes of comparison, "demand schedules" of the type used here may be superior to those that comply with the usual definition of demand. For example, if a seller is interested in patterns of secondary brand preference, exclusive of switching occurring principally on the basis of subjects selecting the lowest-priced item, then the procedure followed in this study would be superior. On the other hand, there will be many instances when a measure of demand in accordance with the traditional definition will be required, and a way around the difficulties associated with increasing the number of buying decisions will have to be found. If data were being gathered to aid in the solution of a specific problem, it would be practical to work with a single classification and obtain the required data without running the risk of excessive fatigue on the part of subjects.

Results

The "demand curves" for the more popular brands included in the study are shown in Figures 1-4. The original price and number of buyers who preferred the brand at that price are indicated by a solid dot adjacent to the letter designating the brand. Prices and quantities are shown through the range of a 5-cent increase and a 5-cent decrease.⁷ The price elasticities for the total change to that point have been computed for increases and decreases in price of 1 cent, 2 cents, etc., through 5 cents

⁷ An estimate of the reliability for points above the original price on a "demand curve" for the student population used in the experiment may be illustrated as follows: Let p be the probability that a person who has been chosen at random from the subpopulation of buyers of Brand A toothpaste will *not* switch to another brand if the price of Brand A is increased by three cents. A point estimate of p , call it $\bar{p}$, is given by the fraction 37/81, the proportion of experimental subjects preferring

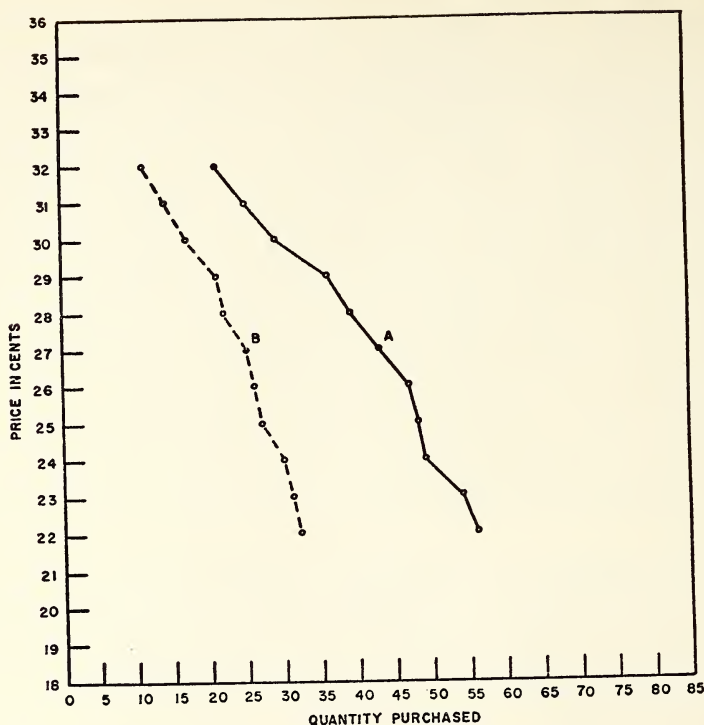


FIGURE 2. Demand curves for cigarette brands preferred by 20 or more subjects.

(Table 1). The elasticities displayed for price reductions are subject to the qualifications outlined in the preceding section and are used exclusively for purposes of comparing one brand to another. Average price elasticities for a classification have also been computed for toothpaste and cigarettes. Since these represent a weighted arithmetic average of the elasticities of all individual brands in the classification, they should not be interpreted as applying to total industry demand.

Although it is impossible to generalize about the behavior of all buyers, it is interesting to review some of the characteristics of demand displayed by those who took part in the simulated shopping trips. For

Brand A who continued to buy it after the price of Brand A had been increased from 31 cents to 34 cents. A 90 percent confidence interval on p is given by the formula:

$$\bar{p} - 1.645 \sqrt{\frac{pq}{n}} \leq p \leq \bar{p} + 1.645 \sqrt{\frac{pq}{n}}$$

For a price increase of 3 cents on Brand A toothpaste, the 90 percent confidence interval is: $0.37 \leq p \leq 0.55$.

A similar procedure could be used for the points below the original price on a "demand curve" by letting p be the probability of choosing a person at random from the subpopulation of buyers of all brands of toothpaste other than Brand A who will switch to Brand A if all brands but the one originally preferred by the person are reduced in price by a given amount.

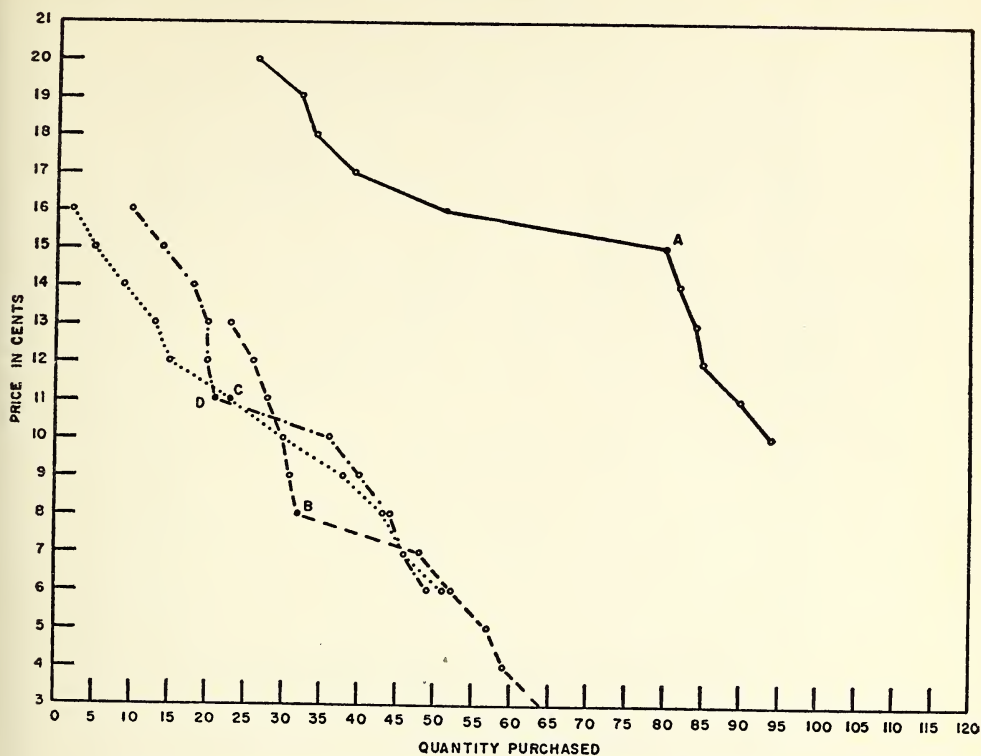


FIGURE 3. Demand curves for toilet soap brands preferred by 20 or more subjects.

simplicity, each classification will be discussed separately. To facilitate identification, the brands in each classification are lettered in alphabetical order beginning with the most popular brand and, in the text, the subscript "t" is used for toothpaste, "c" for cigarettes, "s" for toilet soap, and "h" for headache remedies; thus, for example, Brand B_t designates the second most popular brand of toothpaste.

Toothpaste. The "demand curves" in Figure 1 and the elasticities shown in Table 1 reveal that the subjects varied a good deal in the strength of their brand preferences for toothpaste. For example, Brand A_t, with its large share of the market, gains far less relatively by reducing its price than does Brand E_t, which holds a small share of the market. In addition, Brand E_t loses less relatively by increasing its price than does Brand A_t. The strategic position of a leading brand like Brand A_t will be discussed in the following section, but it is worth note here that the relatively lower price elasticity of Brand E_t under price increases may be attributable in part to the specific medicinal properties of the product, not shared by the other brands in the assortment. Further, when all brands are taken together, a relatively high degree of price elasticity may be observed that remains rather stable over the range in which price was varied.

Cigarettes. The data displayed in Figure 2 and Table 1 show that, for the classification as a whole, brands of cigarettes are less price-elastic

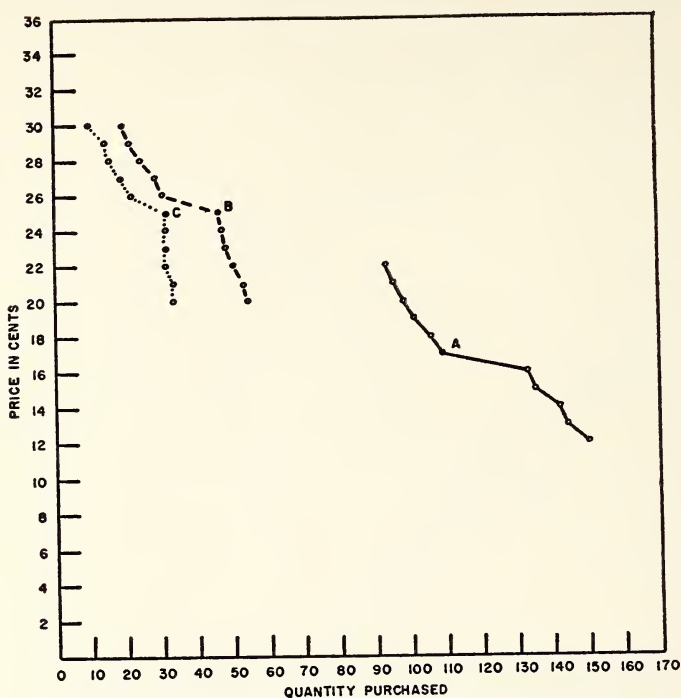


FIGURE 4. Demand curves for headache remedy brands preferred by 20 or more subjects.

than are brands of toothpaste. On the other hand, the elasticities for cigarettes, like those for toothpaste, remain fairly uniform over the range that price varied. Both leading brands of cigarettes, Brand A_c and Brand B_c, gain a good deal less relatively from a reduction in price than do the brands that hold the remaining share of the market. The tendency of leading brands to gain less relatively from a price reduction than do minor brands shows up in both the cigarette and toothpaste classifications. As in the case of the toothpaste, the original prices of all brands of cigarettes were identical. Therefore, to gain an important share of student patronage, a brand had to utilize effectively differential non-price appeals that would prove attractive to a sizable proportion of the buyers. To the degree that a brand was successful in doing so, it reduced the number of buyers who could be easily switched to the brand on the basis of the particular product and promotion appeals it employed. In other words, fewer buyers remain that such a brand could attract because of its past success and the particular form of differentiation it adopted. In other instances, minor brands attracted buyers by somewhat narrower appeals that particularly fill the needs of a smaller group. By concentrating on these special needs, minor brands may make it difficult for the general appeals of the more popular brands to be effective.

TABLE 1
 "PRICE ELASTICITY OF DEMAND" FOR SELECTED BRANDS*

Product	Original Price (Cents)	Change in Price from Original Price (Cents)									
		+5	+4	+3	+2	+1	-1†	-2†	-3†	-4†	-5†
Toothpaste:											
Brand A.....	31	5.0	5.2	5.6	4.4	5.4	1.9	2.3	2.4	2.7	2.6
Brand B.....	31	5.5	5.1	5.8	6.2	3.7	5.0	4.0	4.5	4.0	4.0
Brand C.....	31	5.4	5.6	6.1	7.2	7.2	3.2	4.8	3.7	3.6	4.1
Brand D.....	31	5.3	5.9	4.7	3.8	4.7	5.6	5.6	5.9	5.5	5.9
Brand E.....	31	4.4	3.7	5.5	3.7	3.0	7.4	4.4	5.9	5.5	5.9
All brands‡.....	31	4.9	4.9	5.3	5.1	4.1	4.1	4.1	4.2	4.3	4.5
Cigarettes:											
Brand A.....	27	3.4	3.6	2.9	2.2	2.5	2.5	1.6	1.3	1.7	1.6
Brand B.....	27	3.0	3.0	2.9	2.2	3.2	1.1	1.1	1.8	1.6	1.5
All brands‡.....	27	3.4	3.2	3.8	2.3	3.4	3.7	2.5	3.2	3.7	3.5
Toilet soap:											
Brand A.....	15	2.0	2.3	2.9	3.8	5.4	0.4	0.4	0.3	0.5	0.5
Brand B.....	8	0.7	0.4	0.3	0.3	0.3	4.0	2.5	2.1	1.7	1.7
Brand C.....	11	2.0	2.1	2.2	2.3	3.5	4.0	4.0	3.5	3.0	2.9
Brand D.....	11	1.2	0.9	0.5	0.3	0.5	7.9	5.0	4.0	3.3	2.9
Headache remedy:											
Brand A.....	15	0.4	0.5	0.5	0.6	0.4	3.3	1.8	1.5	1.2	1.1
Brand B.....	25	2.9	3.4	4.0	4.9	8.7	0.5	0.5	0.7	1.0	0.9
Brand C.....	25	3.6	3.4	4.3	5.2	8.9	0.0	0.0	0.0	0.4	0.3

* Coefficient of elasticity = $[p/(p - p')][q' - q]/q]$, where p = original price, p' = altered price, q = quantity demanded at original price, and q' = quantity demanded at altered price.

Basic data from which elasticities of individual brands were computed may be read directly from Figures 1-4. Brands preferred by less than 20 subjects have been excluded.

† See discussion in text of the interpretation of demand that is appropriate for price reductions.

‡ A weighted arithmetic average of the elasticities of all brands in the classification.

Toilet Soap. In contrast to the two preceding classifications, where the original prices of the brands were identical, a distinctly different condition existed in the toilet-soap classification. Here, the leading brand sold regularly at 15 cents, the second most popular brand at 8 cents, and the next two brands in order of popularity sold originally at 11 cents. The kinked demand curve for Brand A_s may be attributed in large measure to the producer's early promotion of a distinct appeal to the consumer's desire for social acceptability. Besides gaining a sizable share of the market through its product and promotion policies, the producer appears to have priced the product to take advantage of the brand's marketing characteristics. If the price of Brand A_s is increased, buyers will be lost in large numbers, but relatively few buyers will be gained if price is reduced by small amounts. On the other hand, Brand B_s holds a very strong price position but a smaller market share. This brand will lose few buyers as a result of small increases in price and, if able to lower its modest price, it can attract substantial numbers of new buyers. Had the toilet-soap assortment found in a typical supermarket been used in the experiment, doubtless Brand B_s would have faced active price competi-

tion. Because such competition was absent, economy-minded subjects were not given an acceptable alternative until the price of Brand B_s increased to a marked degree. In the case of Brands C_s and D_s, both display a high degree of price elasticity when their prices are reduced, but there is a marked difference in the elasticities when their prices are increased. When Brand D_s is compared to Brand C_s, the buyers of Brand D_s show very strong brand loyalty. If the differences in original price are considered, Brand D_s also displays this characteristic when compared to Brands A_s and B_s. A partial explanation of the loyalty evidenced by the buyers of Brand D_s is found in the fact that this soap is used by a number of buyers both as a toilet soap and as a shaving soap.

Headache Remedies. This classification, like the toilet-soap classification, contains items with different original prices. In this instance, however, the lowest-priced item, Brand A_h, has by far the largest share of the market. Like Brand B_s, the lowest-priced toilet soap, it is highly price-inelastic when its price is increased and moderately price-elastic when its price is reduced. Brands B_h and C_h, both higher-priced items, have kinked demand curves characteristic of the differentiated premium product, but in this classification these brands have been unable to attain an important market share. For the group of subjects studied here, the results of product differentiation appear to be less successful for a headache remedy than for a toilet soap. For both headache remedies and toilet soaps the atypical nature of the assortment and of the subjects limits what may be said about the over-all pricing policies of the various brands. In particular, the presence of only one brand in the lower price range creates an unusual assortment for both toilet soap and headache remedies.

Conclusions

Since the experiment reported here was exploratory, the findings cited are necessarily little more than illustrative of some of the kinds of facts about consumer behavior that can be obtained under simulated market conditions. Despite this limitation, it is fair to conclude that controlled experiments offer large promise as a means of attacking a variety of important problems in business and economics that have remained unsolved in the absence of basic facts concerning consumer behavior. For example, it should be practical to employ experimental methods to test the hypothesis that for a given classification there exists a well-defined optimum market share for a single brand. It may also be possible to gain insights into the relative effectiveness of marketing several brands with strong differential appeals, in contrast to expanding the market for a single brand. Further, a better understanding might be developed concerning the degree to which a brand in a given market can expand or protect its market share by price and non-price appeals. The list of similar problems is long, too long to be included here, and the potential applications of research results range from shaping antitrust policy to developing short-range marketing

programs for brand-promoters. The essential point is that the tools of experimental research appear to offer a potentially fruitful means of attacking these problems. By imaginative design and application of experimental methods of research, it is reasonable to expect that the base of facts can be greatly broadened and important progress made toward finding solutions to a number of stubborn problems.

Why Television Commercials Succeed*

.....

HORACE S. SCHWERIN†

TELEVISION ADVERTISING IS ON THE THRESHOLD OF A NEW ERA. WE ARE today seeing the first signs that there soon will be experimentation of a kind and on a scale such as were undreamed of a few years ago. If the trend continues, we can be sure that the commercials of ten years from now will be nothing like the ones we see today.

There are two forces at work to encourage experimentation and change. The first of these is that it has become incredibly costly to remain on television and use an ineffective and imitative sales approach. Advertisers are beginning to realize, in this connection, that relatively little can be done any more in the direction of controlling the cost of reaching viewers, but that a great deal can be done in increasing the effectiveness of the commercials presented to them. Even the shrewdest circulation buy today would give only about a four-to-one advantage over competitors, assuming they made the worst conceivable selection of programs. In contrast, our research reveals that it is possible for commercials for one brand to achieve as much as forty-to-one advantage over another brand in ability to create a preference.

The second force at work encouraging experimentation and change is the development of methods that make it possible for the first time to tell whether new approaches are paying off. The existence of a yardstick like the Competitive Preference technique actually makes it more dangerous for the advertiser to be imitative than to be experimental. The reason for this is that, when his competitor hits upon an imaginative advertising formula, that competitor now has a means of knowing that his new approach will pay off.

* Reprinted from Robert Ferber and Hugh G. Wales (eds.), *Motivation and Market Behavior* (Homewood, Ill.: Richard D. Irwin, Inc., 1958), pp. 321-33. Originally adapted from a talk entitled "Television and Radio Commercials," given at the Marketing Research Conference, University of Michigan, March, 1955.

† Schwerin Research Corporation.

The most important single thing that this new measure has brought out is research proof of something that creative people have long hoped was true: Remembrance of copy points is not the whole answer to commercial effectiveness. In the case of conventional commercials, where convincing demonstration is used to put over the right sales idea, there is a close relationship between remembrance and effectiveness.

But that is only part of the story. There is another area besides convincing demonstration, an area which might be called mood or fantasy, that does not necessarily have much to do at all with putting across explicit sales ideas. A commercial of this nature establishes its own world, within which viewers accept actions and breathe in impressions that they would reject if the mood of the commercial were logical rather than emotive. Commercials of this type have proved extraordinarily effective in swaying viewers toward the brand advertised; and we are receiving more and more commercials of this type to study from advertisers who see which way the wind is blowing.

What has been outlined above might be called "TV's Law of Extremes." In examining the commercials tested that have proved to be most effective, we have found two distinct types. At one end of the curve are commercials where convincing proof of a sales claim is advanced. At the other end of the curve are commercials that create a mood. It is in this second area that boundless opportunities for experiment and progress lie.

Advertisers are basically interested in the cost of effecting an increase in preference; that is, by getting new customers at the lowest possible cost. The way we collectively look at it, there are two distinct areas: First, what does it cost to get into a thousand homes per commercial minute? Second, once a thousand people are captured, how much are the people influenced with commercials, or what is the increase in preference? To provide some perspective on the problem, I would like to cite some costs of getting into a thousand homes.

We use some data borrowed from Mr. Nielsen, based on 98 evening programs on three major networks for the first part of November, 1954, which shows the cost per thousand homes per commercial minute; the best buy was \$1.86 and the poorest, \$7.18, a difference of four to one. We took 95 per cent of the cases in order to get rid of the extremes, where the people are paying excessively high or extremely low rates.

When we consider the second area—what is done once the people are captured—we find an entirely different range. Thus, of commercials studied in 1954, the best commercial had 40 times the effectiveness of the poorest.

The range of efficiency is so startling that, in spite of the fact that circulation and cost are tremendously important, this area seems to be the key to the future in television.

A MEASURING TECHNIQUE

The technique on which the above estimates are based is a laboratory technique.¹ In most of the testing in the greater metropolitan area of New York we attract people from a 40-mile area. We do not go out to them, we invite them into our own theater, located on 46th Street and Sixth Avenue. In Toronto we test at the University theater. In London we plan to test at a small theater on Trafalgar Square. In all areas the techniques are the same.

The people come to the theater. As they walk in, each individual is handed a coupon, divided into two columns, with identical numbers on both sides; one half is retained and the other half dropped into a box.

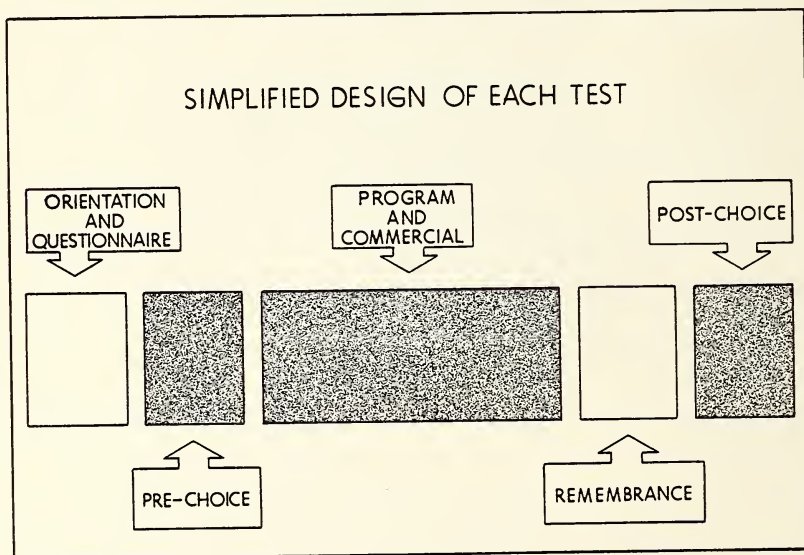


FIGURE 1.

This enables us to identify each individual. We have between 300 and 400 people seated in the theater. Tests are run every night of the week, five nights a week, and three afternoons.

This is schematically what happens: First there is the orientation and the questionnaire (see Figure 1). In the orientation a test director shows the people a series of color slides to get them to relax. He emphasizes the point that they are there to help improve programs. Immediately following that, each individual is led through a rather extensive questionnaire, wherein data are gathered concerning personal and socioeconomic

¹ Credit is due to the companies who gave us financial support and also loaned us their very fine talent from their research departments, namely, General Mills, RCA Victor Corporation, the Borden Company, Toni, and Miles Laboratories of Elkhart, Indiana.

classification as well as regarding market information pertinent to whatever the particular subject matter.

Following the orientation and the filling out of the questionnaire, at the point in time labeled "pre-choice" in Figure 1, the people are told that to reward them for having come, a series of door prizes will be given away. Everyone must check, before the winner is drawn, which of the products he wants sent to his home should he prove to be the winner.

For example, the people may have had a chance of winning a Remington electric shaver, a Schick, Sunbeam, or, if they don't want a razor, \$25.00 in cash. They have had an opportunity to win a year's supply of any of these instant coffees: Chase and Sanborn, George Washington, Maxwell House, or Nescafe. They have had an opportunity to win a year's supply of whichever brand of toothpaste they want. Then the drawing follows. The drawing is very simple. It merely consists of a youngster being asked to come up and pull a number out of the basket. The winner is then identified by the test director. At this point the people still don't know what is to follow in the session. At least they have no basis for knowing either that they are going to be seeing commercials or what the commercials are going to contain.

Following the prechoice, there is a half-hour television show with the commercials in the normal position in the show. Immediately following the viewing of the program, the people are asked to play a little game with us. They are handed a sheet of paper and are asked to give the name of the product advertised and everything they remember having seen or heard about the product in the commercial.

Following that the audience is told that we know there were only a few winners previously and that we would like to reward additional people. At this point, there are additional drawings for prizes. They are given fresh forms identical to the ones they had earlier, and they have another opportunity of winning a series of door prizes. Again they make a choice of any of the electric shavers or the \$25.00 in cash. Again there is a drawing. They make a choice of any of the instant coffees, and then a drawing. They make a choice of any of the toothpastes, then a drawing.

If we were studying a particular commercial within a program we would compare what the people wanted before they saw the program with what they wanted after seeing the program and what they remembered immediately following. This group cannot be used for another commercial effort because they are already preconditioned by what they have seen and by their participation in a test. Therefore, to examine this commercial in relationship to another commercial we must use an entirely separate group of people. To compare the results obtained from two or more groups exposed to different commercials, it is critically important that the groups be identical. The research is no better than our

ability to make the groups precisely the same with regard to the important factors.

How can this be done? The number of sessions run within the past eight years alone exceeded 2,000 groups of people. According to the laws of probability, perfect cross sections coming in by accident would be most unlikely.

The means used for obtaining a cross section is postselection. For example, if an equal number of men and women is required and there are 150 men and 250 women, we eliminate 100 women. Of course, that is gross oversimplification, for it is necessary to match on three or four factors and keep them cellularly separate.

Of all the factors that are critically important, by far the most important is the prechoice of the product. Figure 2 shows three separate

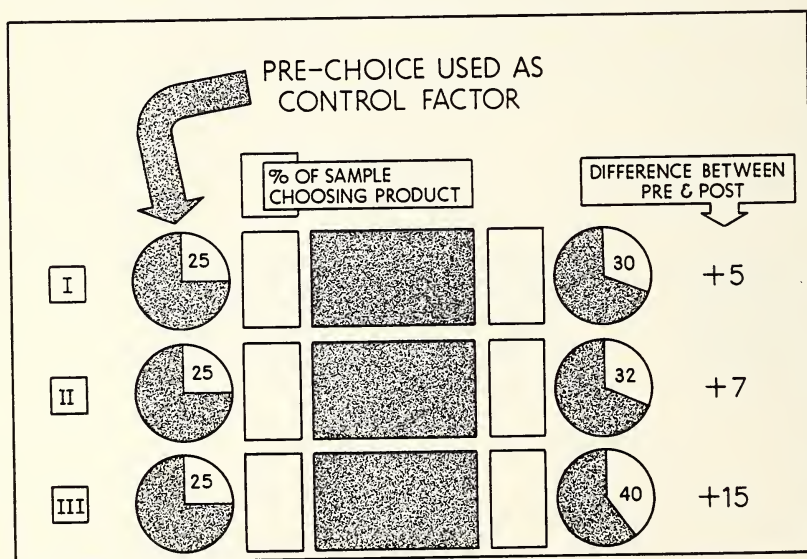


FIGURE 2.

groups matched on choice of a hand lotion. To illustrate, 30 per cent of the women wanted this hand lotion in Group 1. That is reduced to 25 per cent, assuming that is the figure selected, by eliminating 5 per cent of the women who wanted it but maintaining all other factors, keeping them cellularly separate. Of Group 2, 20 per cent wanted the hand lotion. That is brought up to 25 per cent by dropping out the people who wanted other brands until the residue is 25 per cent.

This is a very costly method of operation. It not infrequently requires two or three gatherings of people before the residual sample is adequate. In one instance the process of elimination required working six times in order to get the residual sample adequate for a study of Ken-L Ration because of such big differences between dog owners.

Now we have three different groups at the same level. Group 1 started with 25 per cent of them wanting this brand of lotion before seeing the program. After seeing the commercial in the program, 30 per cent wanted it, an increase of 5 per cent. Group 2, starting with 25 per cent wanting the brand before seeing the commercial, ended with 32 per cent afterwards, an increase of 7 per cent.

Group 3, starting at 25 per cent before the commercial, ended up with 40 per cent, an increase of 15 per cent. We conclude that commercial No. 3 has a stronger influence than commercials 1 or 2, assuming the differences are statistically significant.

RELIABILITY

Let us now examine briefly the reliability of this technique. We are offering the same people at two different points of time the same opportunity of winning a series of products. Do they choose the same when there is no reference to a particular item in between? For example, we offered people cases of Schlitz, Ballantine, Pabst, Miller, Budweiser, Blatz, or \$25.00 in cash. We continued with the regular session, but no reference to beer was made. Then we offered at a second drawing the same list of products or \$25.00 in cash.

In the first instance 21 per cent wanted Schlitz; in the second instance, 22 per cent. Ballantine in the first instance, 17; in the second instance, 17. Pabst in the first instance, 12; in the second instance, 13; Miller in the first instance, 12; in the second instance, 12. Budweiser in the first instance, 10; in the second instance, 11. Blatz in the first instance, 5; in the second, 3. Twenty-five dollars in cash in the first instance, 23; in the second instance, 22.

The changes were not significant. Of course, this had to be done a multiple number of times, as may be illustrated by preference for Schlitz on three different studies on three different groups, wherein there was no commercial in between (see Figure 3). We of course matched to get the same per cent of the people wanting Schlitz, so the groups were identical on brand preference as well as other important characteristics.

In Group 1, 21 per cent wanted Schlitz; no commercial, they ended with 23 per cent. The second group started with 21 per cent wanting Schlitz and ended with 22 per cent wanting Schlitz, no commercial shown. Group 3, 21 per cent, ending with 22 per cent the second time, and no commercial.

We conclude that if we do nothing, nothing happens. This is, of course, an entirely negative type of reliability and evidence is needed on the positive side of reliability. If the same thing is done with groups consistently, something happens.

The same commercial was tested for Toni home permanent waves on three different groups. All groups were matched. At the start, 38 per cent of the women wanted Toni home permanent waves. After the commer-

cial 46 per cent wanted Toni, an increase of 8 per cent. Of the second group, 38 per cent wanted a Toni before the commercial and after the commercial 47 per cent wanted it, an increase of 9 per cent. Of the third group 38 per cent wanted a Toni before the commercial and after seeing the commercial 45 per cent wanted it, an increase of 7 per cent.

Statistically there is no difference between the increments of 7, 8, and 9, so we say that the commercial within the normal limits had the same effect on each of the three groups.

Returning to Schlitz for a moment, we started with the 21 per cent wanting Schlitz. When we did nothing we remained at the same level;

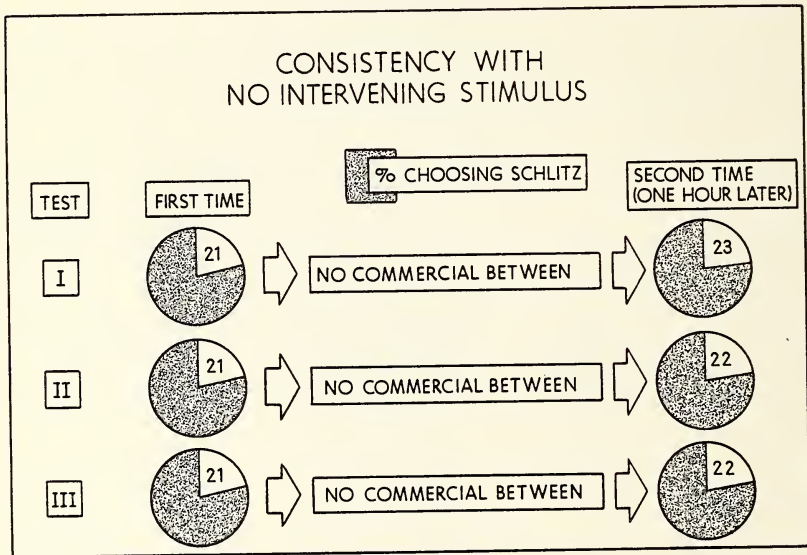


FIGURE 3.

22 per cent wanted Schlitz, or an increase of 1 per cent. On the other hand, if we start another group with 21 per cent wanting Schlitz and they see a commercial for Schlitz and end up with 29 per cent wanting Schlitz, the increase must be attributable to the only change in the experimental structure, which is the inclusion of the commercial.

One might suspect that people would tend to show an increase after seeing the commercials, but here is what actually happens. In the case of a well-known drug product, before the commercial 15 per cent wanted it, and 14 per cent after; no significant change. In the case of a well-known toothpaste, 11 per cent wanted it before and 12 per cent after; no change. Fifty per cent wanted to win a certain very popular, heavy appliance before the commercial; after the commercial 45 per cent wanted it, a loss of 5 per cent. None of it went to cash; it all went to another make. So it is possible actually for commercials to have a boom-crang effect.

A large number of studies have also been conducted in the food field. Of the food product commercials studies, 50 per cent showed no change from prechoice to postchoice that we could measure; that is, the difference from prechoice to postchoice was no greater than if there were no commercial between the prechoice and the postchoice. Some commercials, however, did give increases. Twenty-seven per cent gave an increase of 5 to 10 per cent. Twenty per cent showed an increase from 11 to 20. Three per cent actually gave an increase over 20.

To cite the effects of the commercials that gave increases, the first example is a baking product. Twenty-eight per cent wanted this baking product before the commercial and 32 per cent after, an increase of 4. They were considering a radically new campaign. Before they even tried it on the test market we studied it, with these results: 28 per cent wanted it before and 49 per cent after the commercial, an increase of 21 per cent.

At the time they were running that particular campaign they had a 21 per cent share of the market. Precisely one year later they have a 40 per cent share of the market, without any major change in advertising appropriation.

It is rather interesting to note, in this increase of 21 per cent, that entirely different things are remembered. The campaign that did not do so well had a high remembrance of details, such as the very fine ingredients and the fact that one could bake easily with the ingredients. The campaign that did well had no remembrance of the fine ingredients, no remembrance of anyone being able to bake with ease, speed, and convenience. There was remembrance of good results and the factor of taste.

Let us now consider commercials for television sets. Forty-seven per cent wanted a certain advertised television set in the prechoice and 61 per cent wanted it after the commercial, an increase of 14 per cent. In trying to sell a "clearer picture" idea, this commercial was presented by engineers in white coats, engineers in blue suits, all kinds of professional people, authorities of all types. The campaign that got the highest increase, however, was presented by a baby-sitter. The woman got up and the first thing she said was, "I know nothing about television." She said, "But I do have to earn my living going from home to home. As soon as the folks leave I quickly turn on the TV set. When I see it is this make I know I am going to get a crystal-clear picture." Interestingly enough, 41 out of 100 people remembered "crystal-clear" and the woman who didn't know anything about television, while 13 out of 100 remembered the engineers' statements.

In beauty products, in spite of the fact that many commercials gave no increment, we have had commercials with an increase as high as 40. One product started with 20 and went to 60. In the appliance field quite a few have done nothing, yet some commercials have given as high as 25. In the drug field quite a few also gave no response. The highest has been an 11

per cent gain. In soaps and cleansers, many commercials have given nothing. The highest we have had is 10.

WHY EFFECTIVENESS VARIES

In showing commercials and studying the increments, the highest effectiveness was obtained the more clearly the product was demonstrated, that is, when the public had an opportunity of using their own eyes to see the payoff idea, such as with Hazel Bishop lipstick—it won't smear off (and they show it); it won't stick on your cigarette (and they show it not sticking on the cigarette). Where the demonstration is one in which

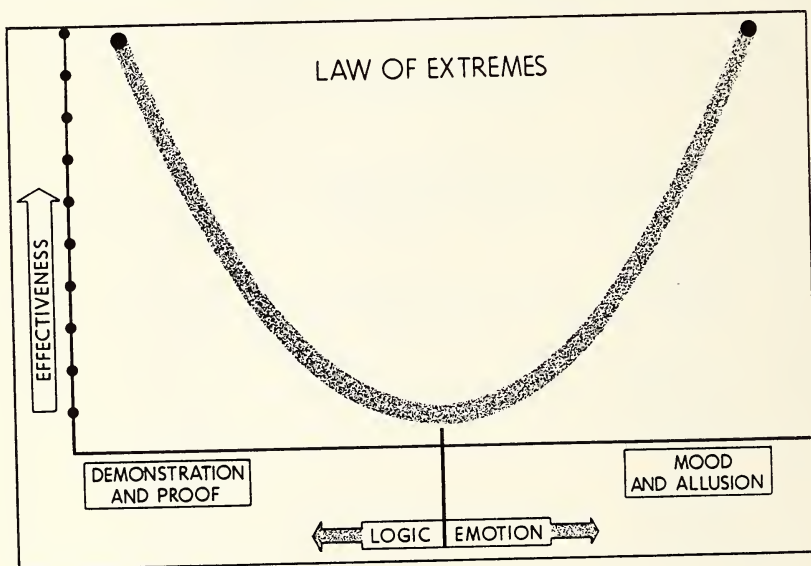


FIGURE 4.

the people are seeing the payoff of the product, invariably, in terms of effectiveness, we are way up on the scale.

With commercials farther down the scale people are no longer being allowed to use their own eyes; somebody is doing it for them. Well-known authorities are saying, "This is the best product." In no instance does that do as well as where people can use their own eyes. Still farther down the scale are commercials in which somebody who is not even an authority stands behind a desk and hammers on the desk, assuring people that his product is best.

We found in one case that the lower the belief, the lower the effectiveness; the higher the belief, the higher the effectiveness. The lower the remembrance, the lower the effectiveness; and the higher the remembrance, the higher the effectiveness. So there are three factors correlated: The higher the remembrance and the higher the belief, the higher the effectiveness, and the difference between the pre- and the post-increment.

Fortunately, the same basic product was presented in a variety of ways. For example, let us take soap. A shampoo was presented in various ways. An actual demonstration was first, showing the payoff idea of the product or the motivating idea of the product by live demonstration.

In a simulated demonstration, a cartoon or some other device is substituted for the live demonstration of the payoff of the product. In an illustration of results, the picture is shown of a fat woman before using the product, and then a picture of a thin woman is shown. That is indicative of what we mean by results. In direct testimony somebody stands up and says, "Trust me. This is the best product in the world." In in-

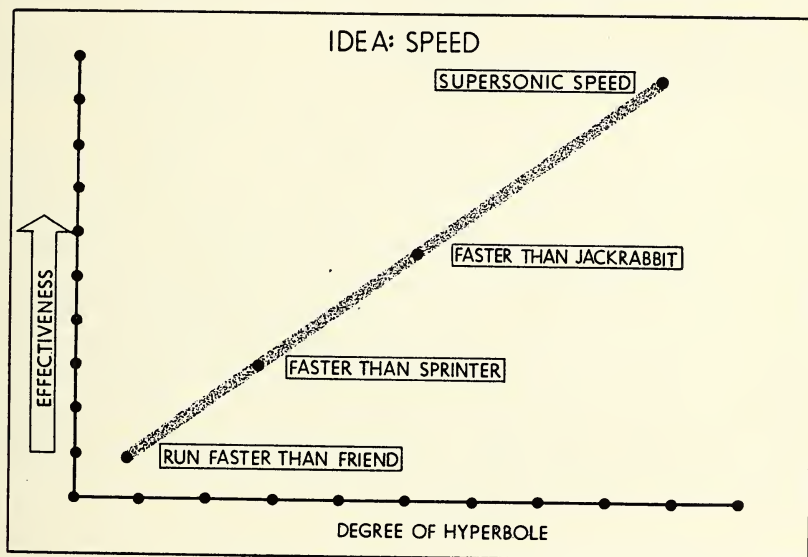


FIGURE 5.

direct testimony, we could not figure out why the commercial was produced.

I would like to point out that where the actual demonstration took place the cluster of the highest range appears. Where commercials tended to go into other approaches they did not do as well.

The demonstration in itself is subject to variables. For instance, one model is shown getting the best possible suds when she uses a certain shampoo. Then there is a model who is making a comparative demonstration. This model uses a common, ordinary shampoo and she doesn't get suds. The difference is rather interesting. Showing the one model alone using the product, we got 20 ideas back to 100 people. When we showed a rival product with two models, we got 40.

Most of the commercials we are talking about (see Figure 4) were essentially radio commercials to which video was added. The visual was not used as a prime means. We then move into another area that nobody

is expected to believe. It is obvious that it is fantasy. The farther we get into the mood and illusion area, the greater the increment. The single highest scoring commercial in terms of effectiveness tested so far had only 30 ideas remembered out of each 100, which is about one-sixth of normal for that product. It had no belief because nobody was expected to believe the particular fantasy shown. The only thing that happened was that women sure wanted the product!

Indication of effectiveness of commercials in this area of emotion is a commercial for sneakers (see Figure 5). The vertical scale is the effectiveness, and the horizontal the degree of hyperbole. The point of one commercial was that the boy can run faster than his friend because he has sneakers on. We get low interest.

As a result of wearing the sneakers he is faster than a sprinter. We get higher interest. When as a result of wearing the sneakers the boy can travel at supersonic speed we get even higher effectiveness.

This may seem silly, yet there are many other examples. Up to very recently most advertising people have been concentrating on getting across more ideas, having to prove them, and having to demonstrate them. For some reason they seem to forget that there is more than one way of influencing people. Some advertising is capable of arousing the desired reaction. The very best commercials—and unfortunately they are very far and few between—are commercials which are in the area of “mood,” or whatever one may want to call it. Most commercials in past years have jumped from logic to emotion. The mood of the program plays a large part, particularly with commercials which are selling via mood, emotion, illusion, allusion, whatever is the correct word.

Field Experimentation

THE FIRST ARTICLE, BY SEYMOUR
Banks, is concerned with the ef-

fect of a new packaging material upon preference and sales of two varieties of bakery goods: layer cake and Danish pastry. Banks is interested in determining the extent to which preference tests (which are usually faster and cheaper than experiments) can be used to replace experiments, especially in situations where relatively minor product features are being examined. In his paper, the experimental results are used as the basis for attempting to validate the use of a preference test as a predictor of the results of a sales test. One feature of his analysis, which is of particular interest given the problem of designing an experiment, is his use of historical sales information as a partial basis for determining what units should be used for measuring the results.

A useful exercise for the reader interested in testing his insight into experimental techniques, would be to contrast the work done by Banks with that done by Gridgeman. Could Gridgeman's design be modified in a fashion that would bring it closer to that used by Banks in order to attempt to validate his results?

Raymond Jessen presents a simplified, hypothetical illustration of the way in which the latin square design can be extended so as to measure not only the effect of varying advertising expenditures, but also the carry-over effects of advertising from one point in time to another. The latin square with modification for estimation of carry-over effects is known as a switch-over design. Jessen's switch-over design provides an illustration of the way in which experimental units can be organized, by area and time period prior to the administration of the treatments, so as to increase the precision of the experiment.

Henderson, Hind, and Brown present a second example of the use of a switch-over design. They report an experiment concerned with the

evaluation of two promotional themes for Washington State apples. One of the themes stressed the various uses of apples, while the other emphasized their healthful qualities. The experiment included variations with respect to the above themes in both in-store promotion and newspaper advertising. As with the Jessen proposal the treatments being tested are balanced by city and by time period. However, in contrast to the Jessen article, considerable use is also made of supplementary information available during the time the treatments were being administered (for example, produce sales, display space for Washington State apples, price of Washington State apples, and so on). The authors use this contemporaneous information in three ways: first, to increase the precision of their estimate of the relative effectiveness of the two campaigns; second, to learn something about the degree of substitutability of Washington State apples for other fruits; and third, to learn something about the effect of other promotional techniques (such as display space) that were not deliberately varied in the experiment.

Belson reports an attempt to evaluate the effect on viewers of a BBC television program, "Bon Voyage," with respect to their knowledge of French words and their attitudes about France. The author argues that the difference between viewers and nonviewers can be looked upon as being made up of two effects: the effect of watching the program and the effect of self-selection that occurs in the process of deciding to watch the program. Belson presents an attempt to use supplementary information with respect to nonviewers with the objective of adjusting this difference for the effect of self-selection, thereby isolating the effect of the program alone.

Though the Belson article was placed in the section on field experimentation, it is not an experiment. The motivation for including this observational study at this point, is similar to that for discussing observational research in "Research Design in Marketing Analysis," elsewhere in this book. The contrast of this report to the previous ones in this section may enable the reader to get a sharper insight as to the characteristics of an experiment and as to the conditions that permit experimentation. This article also acts as a transition to the readings in the section to follow, which reports findings based on observational studies.

The Measurement of the Effect of a New Packaging Material upon Preference and Sales*

SEYMOUR BANKS†

Introduction

SALES TESTS CAN BE USED TO AID A MANUFACTURER IN MAKING CHOICES between alternatives of package design, product formulas, advertising appeals, etc. However, where relatively small changes in product¹ are being considered, sales tests usually will reveal the best of the proposed alternatives only after a long period of time (best is defined in terms of consumer satisfaction rather than physical measurement of product variables). Because of this limitation of sales tests, other types of tests are also used, most of them requiring some statement of preference by individuals. There are two reasons for using preference tests: (1) they give the same information as sales tests but give it faster and more cheaply and (2) they give subjective information—on the reasons for preference—that the data of the sales test cannot yield. The interest of this paper, is focused only on the first reason.

Before tests of preference can be accepted as substitutes for sales tests, it must be shown that the two tests will give substantially the same answer in given circumstances and that the preference tests are really cheaper and faster. An opportunity for an experiment on this methodological issue arose in a test of a new bakery packaging material. Packages made with this new material looked exactly like the old ones but were much more grease- and moisture-proof. The new packaging material was more expensive than the old; however, the increase in cost could

* Reprinted from the *Journal of Business*, Vol. XXIII, No. 2 (April, 1950), pp. 71-80 by permission of The University of Chicago Press. Copyright 1950 by the University of Chicago, all rights reserved.

† Leo Burnett Company, Inc.

¹ By "product" we mean the bundle of utilities which represents the product in the consumer's mind. This includes the physical qualities of the product, its container, and advertising and merchandising factors.

not be passed on to the consumer, since it was only a fraction of a cent a package. Therefore, sales had to be increased to cover the increased cost. Research was undertaken to determine what increase in sales might result from switching to packages made with the new material. The interest of the writer was largely in the methodological problem raised above. Therefore, a pair of matching tests was used in this study: a sales test and a preference test. The results of the two tests were compared for qualitative and quantitative agreement.

The two tests were carried out in the same neighborhood, thus eliminating the difficulties arising from population differences. Two varieties of bakery goods were used, layer cake and Danish pastry, both selling at thirty-five cents.³ The statements of preference were obtained by personal interviews in two localities within Chicago. The sales test was conducted in two sets of stores within each of these neighborhoods. Each of the neighborhoods chosen constituted a delivery route covered by one deliveryman of the wholesale baker co-operating in the study. The stores on each route were divided into two services with delivery made upon alternate days. Preliminary choice of areas was made after comparison of sales of the two routes for equality. Some of the routes of the bakery lay in predominantly residential neighborhoods, other routes delivering to stores in highly industrialized districts. Routes were chosen that lay in residential neighborhoods because it would be much easier to draw a representative sample of the people who shopped in the neighborhood stores in these areas than in the industrialized districts. For the determination of the actual increase of sales due to the introduction of the new package, this procedure is doubtful, since there is no assurance that these are representative neighborhoods.⁴ However, this is not important for the methodological experiment, which was interested only in comparing results obtained by the two types of test in a given environment.

Sales Test

The sales test was an indirect one. It was not possible to put bakery goods in the two types of package side by side in the test stores for consumer choice. Instead there was an indirect comparison of the sales of cakes in the old package in one group of stores against the sales of cakes in the new package in a matching set of stores. It was assumed that the competition faced by Brand X cakes was the same in both sets of stores. Thus, the final results of the tests were in terms of how well Brand X

³ These bakery items were chosen for the following reasons: they offered sharp tests of the qualities of the package; they were good samples of bakery goods; and they represented a fairly large portion of the sales of wholesale bakers.

⁴ A factor omitted in this test and one which would affect the ultimate sales effect of the new package is the discussion of its use in advertising. Even if there is no significant preference for the new package, the advertising can have a new theme which might lead to a permanent improvement in sales.

cakes in both types of package did against their real competition—other brands of bakery goods.

The sales-test procedure was adapted to fit into the regular operations of the wholesale baker who co-operated in the study. The bakery packaged its cakes on a production line, and only one type of package could be used during each day's production. If one attempted to use both packages for a given baking, far too much time would be lost in changing from one package to another and keeping the cakes in the two types of packages separate (remember, externally the packages were identical). The regular delivery procedure was used to prevent store-owners from realizing that a test was in progress and from feeling that some special effort was required from them. The only sales-test design open under these circumstances put one type of package on one service on the two routes on one day and the other type of package on the other service on these routes the next day. After collection of sales data from this arrangement for some time, it was to be reversed. However, the test was ended before this could be done.

Base Period. Collection of sales data for a layer cake and Danish pastry in the old package was made for a period of seven weeks in the spring of 1948. These sales figures were examined to see whether the selection of delivery routes was sound and to obtain base data for the experimental period. Sales data were recorded as net weekly sales per store, the difference between the number of items put into the store fresh and the number taken out as stale. This gave many stores net sales of zero, but they were kept in the sample because they represented a sales opportunity. Stores where no cakes were left were not counted in the sample. The number of stores at which cakes were left for sale fluctuated widely from week to week. The experimenter had no control over this.

The basic tabulation of sales for each variety during the base period showed considerable variation in number of stores in the sample from week to week within the services of the two routes. The first consolidation of the data was to convert it into average net weekly sales per store for each service (combining both routes). This was done for both layer cakes and Danish pastry (Table 1).

The averages of net weekly sales per store show wide variation from week to week within services and between services. For layer cakes the difference between the two service means for the base period is significant at the 5 percent level of confidence. The difference between the base period service means for Danish pastry is not significant.

The question arises whether use of simple over-all averages is sufficient to reveal the underlying forces which contribute to the fluctuation of the recorded sales data. Variation in sales can arise from week to week fluctuations of differences in sales between routes, differences between

TABLE 1

AVERAGE NET WEEKLY SALES OF CAKES ON SERVICES A AND B, MARCH 22-MAY 8, 1948

Date	Layer Cake				Danish Pastry			
	Service A		Service B		Service A		Service B	
	$\bar{x}$	$\sigma_{\bar{x}}$	$\bar{x}$	$\sigma_{\bar{x}}$	$\bar{x}$	$\sigma_{\bar{x}}$	$\bar{x}$	$\sigma_{\bar{x}}$
March 22-27.....	1.950	(0.179)	2.804	(0.164)	2.514	(0.412)	3.045	(0.398)
March 29-April 3.....	2.374	(0.171)	2.089	(0.227)	3.054	(0.391)	2.244	(0.278)
April 5-10.....	2.509	(0.179)	2.247	(0.180)	2.867	(0.361)	2.848	(0.372)
April 12-17.....	2.614	(0.194)	2.918	(0.208)	2.432	(0.261)	3.049	(0.418)
April 19-24.....	1.802	(0.246)	2.339	(0.170)	2.031	(0.372)	1.969	(0.269)
April 26-May 1.....	2.502	(0.159)	2.732	(0.029)	3.194	(0.402)	2.739	(0.370)
May 3-8.....	2.253	(0.130)	2.108	(0.163)	2.952	(0.380)	2.400	(0.372)
Overall.....	2.297	(0.065)	2.499	(0.065)	2.763	(0.102)	2.635	(0.102)
$\bar{x}_A - \bar{x}_B/\sigma_{\bar{x}A} - \bar{x}_B$	2.20				0.88			

service averages within each route, and, finally, differences between stores on each route of each service. The technique of analysis of variance was used to determine the effect of each of these sources of variation in the data.

Analysis of variance depends upon certain requirements which were not met by the data, equal variance in arrays, and normality of observations. In order to make use of the exact tests of the analysis of variance, the sales data were transformed by a modified square-root transformation suggested by Bartlett.⁵ One further manipulation of the data was required for performance of the desired analyses of variance. The usual procedures of calculation depend upon equality or proportionality of numbers of observations in the cells of the tabulated data. The actual sales data had disproportionate numbers of observations because of the irregular variation of stores in the sample from week to week. No methods of dealing with disproportionality of numbers of observations in a three-way table were available. Therefore, the number of observations in each cell of the data were equalized by numbering each observation and drawing the required frequency from a table of random numbers—15 for the layer-cake data and 12 for the Danish pastry sales data. The resulting sales data for each variety, transformed and all frequencies equalized, were analyzed as suggested by Snedecor⁶ (Tables 2 and 3).

The analyses of variance summarized in Tables 2 and 3 were performed to test the significance of the difference between service mean sales for the two varieties of bakery goods after the effects of other

⁵ M. S. Bartlett, "The Use of Transformations," *Biometrics*, Vol. III (March, 1947), pp. 39-51. For a situation in which means are in the range 2 to 10 and zeros occur in the observed data, Bartlett recommends the use of the transformation $\sqrt{x_i + \frac{1}{2}}$.

⁶ G. W. Snedecor, *Statistical Methods* (4th ed.; Ames, Iowa: Iowa State College Press, 1946), pp. 304-9.

TABLE 2
SUMMARY TABLE OF ANALYSIS OF VARIANCE
OF TRANSFORMED LAYER-CAKE DATA
(Number of Observations Equalized)

<i>Effects</i>	<i>Sum of Squares</i>	<i>Degrees of Freedom</i>	<i>Mean Square</i>
Routes.....	2.02	1	2.02*
Services.....	0.13	1	0.13
Weeks.....	3.49	6	0.582†
R × S.....	2.04	1	2.04*
R × W.....	1.23	6	0.205
S × W.....	0.87	6	0.145
R × S × W.....	1.24	6	0.207
Individual stores...	82.99	391	0.2123
Total.....	94.01	419	...

* *F* ratio significant at 1 percent level of confidence.

† *F* ratio significant at 5 percent level of confidence.

sources of variation were removed. Both tables show the same result—the differences between service means were due to sampling fluctuations alone. Therefore, the selection of the routes and services used in the base period test was upheld. Table 2 suggests that the significant variation in service means found in Table 1 for layer cakes was due to the significant variation in route means and route × service interaction. Interaction arises when the effect of one group of objects upon another is not consistent throughout but depends upon the individual members involved. Thus, the average sales of layer cakes in Service A of Route 1 were lower than on Service B, but on Route 2 average sales on Service A were much higher than on Service B.

The significance of the variance attributable to the effect of the different weeks found in both Tables 2 and 3 is not important for the sales

TABLE 3
SUMMARY TABLE OF ANALYSIS OF TRANS-
FORMED DANISH PASTRY DATA
(Number of Observations Equalized)

<i>Effects</i>	<i>Sum of Squares</i>	<i>Degrees of Freedom</i>	<i>Mean Square</i>
Routes.....	0.17	1	0.17
Services.....	0.29	1	0.29
Weeks.....	3.91	6	0.652*
R × S.....	0.03	1	0.03
R × W.....	0.87	6	0.432†
S × W.....	2.51	6	0.418†
R × S × W.....	0.65	6	0.108
Individual stores...	51.66	307	0.1683
Total.....	60.09	335	...

* *F* ratio significant at the 1 percent level of confidence.

† *F* ratios significant at the 5 percent level of confidence.

test, since it was planned to accumulate data over many weeks and since the data to be compared were to be over-all means.

Test Period. The actual sales test was started in June, 1948, and ran eight weeks. It was run exactly as the preliminary test had been conducted except that cakes distributed to one service of the two routes were in the new package and cakes distributed to the other service were in old packages. After eight weeks the sales test was stopped by the bakery because of internal operating problems. Attempts to get the bakery to resume the test were made, but these failed.

Although the sales test was tremendously foreshortened, the available data were analyzed. As before, the average net weekly sales data were first inspected (Table 4). Again, there is considerable variation in aver-

TABLE 4
AVERAGE NET WEEKLY SALES OF CAKES IN OLD AND NEW PACKAGES
JUNE 14-AUGUST 7, 1948

Date	Layer Cake				Danish Pastry			
	New Packages		Old Packages		New Packages		Old Packages	
	$\bar{x}$	$\sigma_{\bar{x}}$	$\bar{x}$	$\sigma_{\bar{x}}$	$\bar{x}$	$\sigma_{\bar{x}}$	$\bar{x}$	$\sigma_{\bar{x}}$
June 14-19.....	1.884	(0.141)	2.146	(0.161)	2.432	(0.294)	2.795	(0.261)
June 21-26.....	2.198	(0.126)	2.161	(0.164)	2.500	(0.264)	2.600	(0.228)
June 28-July 3.....	2.229	(0.152)	2.010	(0.141)	1.893	(0.226)	2.675	(0.236)
July 5-10.....	1.769	(0.119)	2.065	(0.115)	2.275	(0.225)	2.837	(0.272)
July 12-17.....	1.922	(0.119)	1.916	(0.134)	1.886	(0.194)	1.510	(0.186)
July 19-24.....	1.596	(0.130)	1.761	(0.103)	1.959	(0.150)	1.527	(0.194)
July 26-31.....	2.000	(0.116)	2.343	(0.144)	1.415	(0.109)	1.750	(0.175)
August 2-6.....	2.143	(0.172)	1.859	(0.116)	1.605	(0.189)	1.641	(0.135)
Overall mean.....	1.917	(0.049)	2.035	(0.049)	2.006	(0.080)	2.135	(0.079)
$\bar{X}_N - \bar{X}_0 / \sigma_{\bar{X}_N - \bar{X}_0}$			1.68				1.13	

age sales for the bakery goods in the two types of package. Considering the over-all averages, the difference in sales attributable to the new package is not significant for either layer cake or Danish pastry.

This procedure of testing the significance of the difference between service means assumes that the average sales for the two services were equal before the introduction of the new package. However, differences between service means existed during the preliminary period, even though they were found to be nonsignificant after removal of other causes of variation. Another means of determining the effect of the new package is to see if the differences between service means found in the preliminary period were materially changed during the sales-test period (Table 5).

The basic test of this second procedure is the significance of the difference between a pair of differences. This is not so sensitive a test as one dealing with the difference between means. The standard error of a dif-

TABLE 5

COMPARISON OF SPREADS BETWEEN AVERAGE NET WEEKLY SALES STORE BEFORE AND AFTER INTRODUCTION OF LAMINATED PACKAGES

Service	Base Period			Sales Test Period			Δ
	<i>N</i>	Mean	$\sigma^2_{\bar{x}}$	<i>N</i>	Mean	$\sigma^2_{\bar{x}}$	$(x_A - x_B)_{B\bar{P}}(x_A - x_B)_{TP}$
Layer cakes:							
A*.....	670	2.297	0.004225	820	1.917	0.002432	
B.....	623	2.499	0.004225	768	2.035	0.002498	
Diff.							
$(\bar{x}_A - \bar{x}_B)$...		-0.202			-0.118		-0.084
$\sigma_{\bar{x}_A} - \sigma_{\bar{x}_B}$..		0.0919			0.0701		0.1151
<i>t</i>		2.20			1.68		0.87
Danish pastry:							
A*.....	262	2.763	0.01040	341	2.006	0.006314	
B.....	299	2.636	0.01040	363	2.135	0.006262	
Diff.							
$(x_A - x_B)$...		0.127			-0.129		0.256
$\sigma_{\bar{x}_A} - \sigma_{\bar{x}_B}$..		0.144			0.112		0.183
<i>t</i>		0.88			1.15		1.40

* This service carried the new package during the test period.

ference between two means is based upon two contributions of sampling error, while the standard error of the change of the difference between two pairs of means is based upon four sources of sampling error. However, a test of the significance of the difference between a pair of differences can be used in cases where the original pair of means are not equal and the more sensitive test cannot.

For layer cakes the difference in service means during the base period was .202 cake per store per week. The new package was used on the service with the smallest mean sales per store. The previous spread was reduced to .118 cake during the test period, but this reduction is insignificant. For Danish pastry the base period spread was reversed in sign by the test-period sales. Service A was ahead in sales during the base period, but when the new package was placed upon this service, Service B had higher average sales. However, this apparent reversal in sales was not significant.

On the strength of these two tests (Tables 4 and 5), it is concluded that the data of the eight weeks' sales test show that the new package did not increase the sales of layer cake or Danish pastry significantly. The sensitivity of the sales test as a detector of improvement in sales can be gauged from the fact that the minimum increase in sales detectable at the 5 percent level of confidence is .189 layer cake or .301 Danish pastry per store per week.⁷ In terms of the average sales of cakes in the old pack-

⁷ These figures are 1.64 times the corresponding standard error of the change in spread, since this is a one-sided test concerned with both the magnitude and the direction of change.

ages during the test period, these just-significant differences are 9.3 percent of the sales of layer cake and 14.05 percent of the sales of Danish pastry. Increases in sales as great as this or greater would be detected as significant by the test procedures.

Consumer Survey

The consumer survey test studied preferences for the appearance and flavor of cakes packaged in the new and old materials by the method of paired comparison. The research also studied the effect of shelf life upon these preferences—was the pattern of preference between packages affected by the length of time cakes remained in them? Half of the interviews used fresh cakes and the other half of the interviews used cakes four and five days old. This period covers the usual length of time cakes would remain upon store shelves.

As before, it was necessary to fit the consumer survey procedure into the normal operating routine of the bakery. Because the two types of packages were used on alternate days, the survey design overcame this problem by giving each of the packages alternating days' advantages in freshness. These alternating advantages nullified one another, and the final results are not affected by this process. Normal handling of packages used in interviewing was insured by having them delivered to designated stores in the services as part of the regular delivery by the routemen. The interviewers picked up the cakes at these stores and either used them for that day's interviewing or kept them until they were the prescribed age. The fresh cakes used for one-half of the interviews were one and two days old; the cakes used for the remaining interviews were four and five days old (the age does not include the day of baking).

During the interview a respondent was shown two cakes of each variety of bakery goods for determination of preference upon appearance and was given slices from two other cakes of each variety for flavor tests. In these pairs, one cake was in the new package and one in the old. The interviewer was instructed not to tell the respondent the purpose of the test or the difference between the packaged cakes.

The sample used was drawn from maps of the area covered by the two delivery routes. The blocks were numbered, and a systematic sample drawn from each area after a random start. For each block, an interviewer's block location sheet was made up, locating it, giving the corner from which interviewing was to start (this was rotated from sheet to sheet) and the numbered families to be interviewed. The interviewer was required to count families on the block in a prescribed manner, interviewing the designated families. The numbering for each block was done with a constant interval after a random start. Plans called for an average of six interviews per block. The interviewers were instructed not to make call-backs if the designated families were not at home but proceed to the next-numbered family. The time of each successful interview was re-

corded. If an effect of time upon preference appeared, it was hoped that the time distribution of interviews could be used to remove this source of bias. The results of the interviews are summarized in Table 6.

If the difference in package did not affect preferences, the figures in this table would differ from 50 per cent only within a calculable sampling fluctuation. Two of the relative proportions of preference for the new package of the survey are significantly different from 50 per cent: that for the appearance of Danish pastry and that for the flavor of layer cake. The size of the sample and the block-to-block variation in preference are such that the minimum difference in preference detectable by this test is about 10 per cent.

It was found that there was no consistent pattern of change of preference with age of bakery goods. The significant preference for the ap-

TABLE 6
PROPORTIONS OF RESPONDENTS PREFERRING BAKERY GOODS IN NEW PACKAGES

	<i>Layer Cake</i>			<i>Danish Pastry</i>		
	<i>N</i>	%	σ_p^*	<i>N</i>	%	σ_p^*
Appearance.....	230	49.6	2.98	230	67.4†	7.28
Flavor.....	228	65.4†	4.11	229	59.9	5.23

* Standard errors are based upon the sampling design of subsampling within sample blocks.

† Percentages significantly different from 50 at the 1 percent level of significance.

pearance of Danish pastry in the new package appeared the day after baking and continued at its high level. The preference for flavor of layer cakes in the new package was much stronger at five days of age than it was at one. All the other proportions showed only sampling fluctuations around 50 per cent. A tabulation of the preference data in terms of the day's advantage in freshness given alternately to the two kinds of packages showed this to have no significant effect upon preference at either the earlier or the later age levels.

The preference test procedure used cakes whose ages corresponded only to the freshest and oldest cakes on sale in stores. This was done to get the sharpest test possible of the effect of age upon preference for bakery goods in the two types of packages. The findings showed that age had no consistent effect upon preference; therefore, all the preference data were combined and analyzed as if they were based upon bakery goods of the same age. If it had been found that age of cakes used in the preference test should have matched the age distribution of cakes found on sales in stores.

It had been assumed that the time of interviewing had no effect upon preference. This hypothesis was not tested for all preference distributions; but a study of the time distribution of the preference for the appearance of layer cake in the new package showed no time effect. A

chi-square test was set up using as expected values for each hour of the day the over-all proportion of preference for the new package. The value of chi-square obtained from the two distributions was insignificant;⁸ therefore, the original assumption was accepted and the data used without adjustment.

Comparison of Results of Survey and Sales Tests

The methodological problem raised at the beginning of this paper was whether preference and sales tests gave the same answer in a given situation. Comparison of the results of these two types of tests shows inconclusive evidence of qualitative agreement between the tests.⁹ For layer cakes, the survey and sales tests agree that there is no significant preference for either package. For Danish pastry, the two tests disagree. The survey showed a significant preference for the new package— 67.4 ± 14.6 per cent of the respondents preferred the appearance of the Danish pastries in the new package to those in the old. The sales test showed no difference in sales between the two types of packages. The sales test was capable of detecting an increase in sales of 14 per cent between the two types of packages. If the difference in preference indicated by the survey had a direct counterpart effect in sales, the sales test should have revealed it. Instead the sales test merely showed no difference.

There are several reasons why the sales test and the survey test disagreed on the effect of the new packaging material; these reasons spring from the different situations studied by the two tests. In the survey the respondent was confronted with only two packages of Brand X pastry and asked to choose between them on the basis of some single attribute. In the store, consumers did not choose between two packages of Brand X but among a given package of Brand X and other brands and kinds of bakery goods.¹⁰ Choice was based upon a multiplicity of factors. Preference upon package appearance is only one of the many factors affecting purchase of bakery goods. To predict the effect of a given difference of preference upon sales, it is necessary to know the importance of the factor (upon which the preference judgments have been expressed) in consumers' purchase decisions. The data can be used to set

⁸ Chi-square = 3.818; $P (x^2 = 3.818, df = 6) 0.80 \gg 0.70$.

⁹ The following discussion is based only upon the survey findings on preferences for the appearance of cake and packages. Appearance is an important factor in the purchase of this type of cake from the shelves of a grocery store. Flavor differences are slower in their effect upon purchase, since they are not directly competitive at the point of purchase.

¹⁰ The original design of the sales test called for putting both types of packages on sale in each of the test stores. This procedure would have tested each package against each other and against the competition. However, the need to fit the test into the normal operating routine of the bakery ruled out this procedure, and the one actually used was adopted because of its convenience.

a limiting value of the effect of preference upon appearance of bakery goods and package upon sales in this case. The sales test could detect an increase of sales of 14 per cent. The difference in preference found by the survey was 17.4 per cent. Preference upon appearance would have to have a weight of at least 80 per cent in the purchase decision for the significant difference in preference to bring about a detectable increase in sales ($17.4 \times .80 = 14$). Eighty per cent is an extremely high value for the importance of any product attribute in a purchase decision.

Concluding Comment

It is obvious that the situations studied by the preference and sales tests described above were different. It is a truism that the two test situations should be alike for the greatest validity of comparisons of their results. But preference tests based upon subjective response to stimuli can never duplicate purchase situations completely—at least one factor, payment, will be missing and the rest of the test situation simplified. (This is why we do preference tests; imagine reproducing a supermarket in someone's home!) The purchase situation is much more complex than the preference situation. The inevitable result of this complexity is to reduce the importance of any one product attribute or variable in the purchase situation. There will be no simple one-to-one relationship between preference on a product attribute and the sales of the product. The result of this line of reasoning is the conclusion that the choices made by preference tests should be considered as ordinal in nature. A preference test can select the best of a series of proposed alternatives, but it cannot tell how much effect the use of this best alternative will have upon sales.

The question raised at the beginning of this paper was: "Can preference tests be used in place of sales tests for the selection of the best of several proposed alternatives of a product attribute or variable?" It is suggested that the answer to this depends upon the cost factors involved in the adoption of these product changes. When one is dealing with a case where the attribute alternatives are of equal cost or where a new package or ingredient change would bring about a reduction in cost, ordinal choices are acceptable. However, in cases where a proposed change means an increase of cost, cardinal data are needed to answer the question: "Will the increase in sales cover the increase in cost?" For these situations sales tests would be used. Even here, preference tests may be used as filters for the sales test, if many alternatives are proposed, allowing only the best one or two to go into a sales test against the present package or formula.

A Switch-Over Experimental Design to Measure Advertising Effect*

.....

R. J. JESSEN†

AN IMPORTANT GOAL IN DETERMINING THE SIZE OF AN ADVERTISING budget is to maximize profits. To achieve this goal, even approximately, is a very complex problem. Many procedures or rules have been adopted to provide what appears to be a reasonable answer from one point of view or another, but which, most people in the business would agree, are very crude at best. Hence, such approaches as "percentage-of-sales," "all-you-can-afford," "objective-and-task" and "competitive-parity" are used because "scientific" approaches either don't exist, or have not been able to prove their superiority.

The difficulties of devising a scientific approach are fairly well known. The multitude of factors involved and the scarcity of relevant and accurate data are only two of many which could be listed. But work is being done to develop good methods and, with the increasing interest that researchers are paying to the problem, there is no doubt that real progress is being and will be made.

The purpose of this paper is to present a scheme for determining the sales response to varying advertising outlays for a very simple case. It is hoped that the method can be extended to more complex and therefore more realistic cases once the underlying principles are thoroughly understood.

Although we are concerned here with the response of sales to different amounts of advertising outlay, it should be relatively straightforward, at least in principle, to proceed to the determination of the optimum outlay.

Suppose we represent the general relationships existing between the advertising outlay of a firm, its consequent sales and profits as those

* Reprinted from the *Journal of Advertising Research*, Vol. I, No. 3 (March, 1961), pp. 15-22. Copyright 1961 by the Advertising Research Foundation, Inc., 3 East 54th Street, New York 22, New York.

† University of California at Los Angeles and C-E-I-R, Inc.

shown in Figure 1. To keep our model simple, let us assume that the "best" use is made of the advertising budget at each outlay level and that disturbances such as seasonal variations, general state of the market and the economy, competitors' activities, etc., are held constant.

Sales are expected to increase, but not in proportion to advertising expenditures. Increasingly heavy doses of advertising are expected sooner

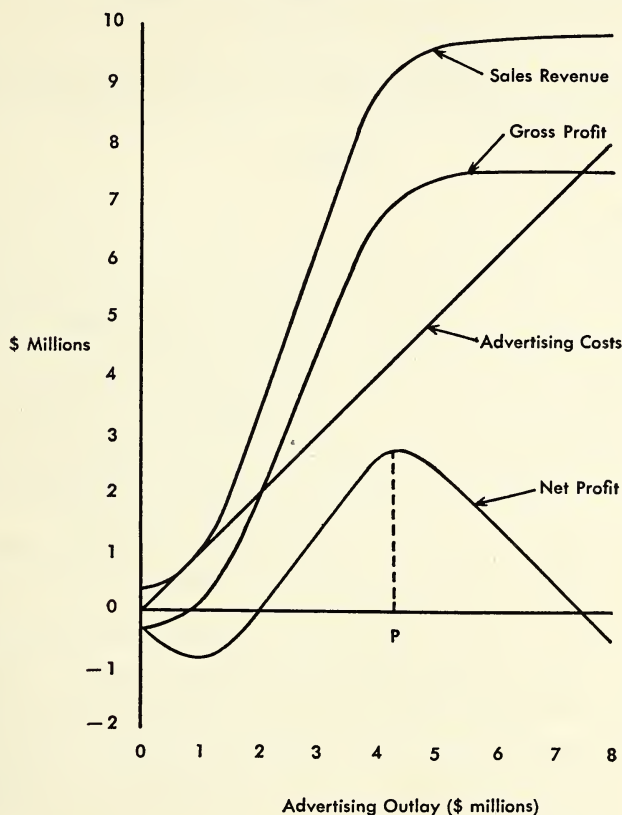


FIGURE 1. General relationship of revenue and profits to advertising outlays.

or later to meet with a diminishing increase in sales. If gross profits are defined as simply Sales Revenue less all costs except those for advertising, we obtain the Gross Profit curve (which may be negative for low advertising outlays). Next, a curve is shown representing increasing Advertising Costs, and lastly, the Net Profit curve which is simply the Gross Profit less Advertising Costs. Net Profit is a maximum at Point "P" where \$2.75 million net profit is made with an advertising budget of \$4.25 million. Any other size of budget will yield less profit.

This paper is concerned with a procedure for determining the top curve—the response of sales to total advertising outlays. The use of the

budget among the different media and the particular kind of copy presented need not be fixed. The conditions which must be satisfied will be mentioned later.

The remainder of the paper will present:

1. A brief statement on basic statistical requirements for planning and conducting experiments.
2. Special problems in dealing with advertising budgets.
3. A layout for a switch-over type design.
4. Some methods of estimating "carry-over" effects.
5. The results of a hypothetical experiment which estimate these effects.
6. Discussion of the applicability of this experimental design.

METHOD

Experiments

A controlled experiment is an investigation in which the experimenter chooses *which* treatment to apply *when* and *where*, and then observes the response. The design and analysis of complex experiments is a branch of applied mathematics largely originating in the biological sciences and in recent years is gradually appearing in a number of other sciences. Its penetration into economics and studies of human behavior has been particularly slow—partly because of the complexity of many of the problems of these fields, and partly because of unfamiliarity with modern developments in modern experimental techniques.

It was a long-held scientific doctrine that good experimental procedure was to study one variable at a time holding all others constant. Thus, one might study the relationship of the occurrence of stomach ulcers in monkeys to kinds, intensity and frequency of stress by varying the intensity while holding kind and frequency constant. Having determined this relationship, one would vary frequency while holding kind and intensity constant, and so on. Such a procedure is not only quite wasteful of time and resources, but provides insufficient information. If there is an interaction between intensity and frequency, it would never be discovered by this method. The correct method is to determine the effects of intensity, frequency and kind, and their interactions *simultaneously* in a single well-planned experiment.

To design experiments appropriate for a given investigation requires some familiarity with experimental design as an art and science, and some of the relevant lore of the field in which it is being applied. If the problem is such that standard techniques, even with modification, will not apply, it may be advisable to devise new procedures.

Advertising Budgets

There are a number of special problems that immediately become apparent when one considers good experimental procedure for measuring the effects of advertising budgets. Often these problems have made some

would-be experimenters shy away with the remark that experimentation may be all right for agriculture in testing for yield effects from different fertilizer treatments or in engineering research, but it just can't be used in the social sciences.

Listed below are some of these problems, with comments:

1. What is a "dose" of advertising expenditure? The amount of dose can be varied by changes in frequency, coverage, space units (e.g., color versus black and white), copy, etc.

2. Carry-over effects. After an advertising treatment is removed, there is still a residual effect left which will continue to influence sales for awhile.

3. Seasonal effects. Sales may have a pronounced seasonal pattern and it may be necessary to remove this influence. Also, a given dose of advertising given in the spring may bring a response different from the same dose given in the fall.

4. Geographical effects. Responses to advertising differ from one locality to another due to income structure, influence of competitors' position, etc.

5. Competitors' programs. If a competitor changes his advertising during the course of an experiment it may confound results.

6. Shifts in consumers' behavior. If the experiment takes a long time to complete, basic consumer behavior in regard to advertising may change.

7. Other factors affecting response. Changes in weather, the world political and economic situation, etc., during an experiment may cause trouble.

8. Measuring sales. Records on sales may not be available for experimental markets or for the time periods desired. Perhaps there have been price changes in the product, or a new premium scheme has been launched.

9. Interference of the experiment with going operations.

A properly designed experiment would have to take problems of the sort listed above into account or else it is doomed to failure.

A Simple Switch-Over Design

Consider the following illustrative problem. A company has sales outlets in 85 market areas into which it divided the U.S. (to make subsequent uses here simple), and its advertising is carried out locally in each of its market areas. Suppose its advertising budget in each market area is now \$1,000 per month. What would be the sales response to budgets of \$2,000 per month? To \$3,000 per month? These outlays would correspond roughly to annual expenditures of one, two and three millions (see Figure 1).

Denote test areas by 1, 2, 3, etc., time periods by I, II, III and budget levels by A, B, C, where the lower budget rate (A) is \$1,000, the medium (B) is \$2,000 and the high (C) is \$3,000.

We wish to put the treatments A, B and C on *each* area during the experimental period in order to eliminate area differences in our comparisons. It is reasonable to believe that the *order* of treatments will affect results; that is, the sales response in each period to an advertising sequence of ABC will likely be different from that of CBA. (This would mean a \$1,000-\$2,000-\$3,000 sequence as contrasted with a \$3,000-\$2,000-\$1,000 sequence.) Since there is a total of six possible se-

quences, a minimum of six areas will be required for the experiment. These can be arranged into two groups of three areas each, the areas having similar seasonal patterns being put together into some group. With this design, each group can be regarded as a separate experiment

		Area		
		1	2	3
Period	I	A	B	C
	II	B	C	A
	III	C	A	B

		Area		
		4	5	6
Period	I	A	B	C
	II	C	A	B
	III	B	C	A

17

FIGURE 2. 3 x 3 Latin square and its complement.

known as a 3 x 3 latin square. There will be two basic types to cover all possible sequences (see Figure 2).

For our purposes it may be better to combine the two squares into a

single experiment with rows (periods) regarded as constant throughout the U.S. In this case the breakdown of degrees of freedom for the purpose of determining "experimental error" would be as follows:

<i>Source of Variation</i>	<i>Degrees of Freedom</i>
Treatments.....	2
Periods.....	2
Sequences.....	5
Residuals.....	2
Error.....	6
Total.....	17

In a later section, suggestions are given for the appropriate method of computing the mean squares of this table for this design.

Carry-Over Effects

The switch-over type design can be useful in eliminating errors due to area differences, both in the average main effects of the budget treatments and in the average seasonal effects, by balancing one effect against another. If carry-over exists, the unadjusted observed effects will be biased. Adjustments will be required if estimates free of bias are to be made of what the sales effect for various budgets would be if they were carried out for extended periods of time.

The nature of carry-over effects in the proposed experiment may be illustrated by extending our example. Suppose our company, whose advertising sales-profit functions are shown in Figure 1, carried out an advertising budget experiment on a set of six test market areas chosen by some randomized scheme. A switch-over type design was used where three levels of budgets were run in the six possible sequences shown in Figure 2. A response to a given budget level is read off the response function in Figure 1. For example, if the budget in a given market area is \$1,000 per month, it would correspond to a national annual level of \$1,000,000 and would produce an expected response of \$1,000,000 in sales at the national annual level, or \$1,000 per month in the individual market area. Likewise, budgets of \$2,000 and \$3,000 per month would produce expected sales responses of \$3,400 and \$6,300 per month, respectively.

In Figure 3, the response to each budget is shown for the two cases: (1) with and (2) without response lag or carry-over. The broken lines indicate the response for each period if there were no lag; the small circles (connected by solid lines) the average response for the period considering that some sales may lag somewhat behind the input of advertising effort. It is assumed that experimental periods for a trial run at a given level are so short that some sales triggered by advertising in one period occur during the subsequent period when a new level is being put to test. It is being presumed that the advantages of a short period are

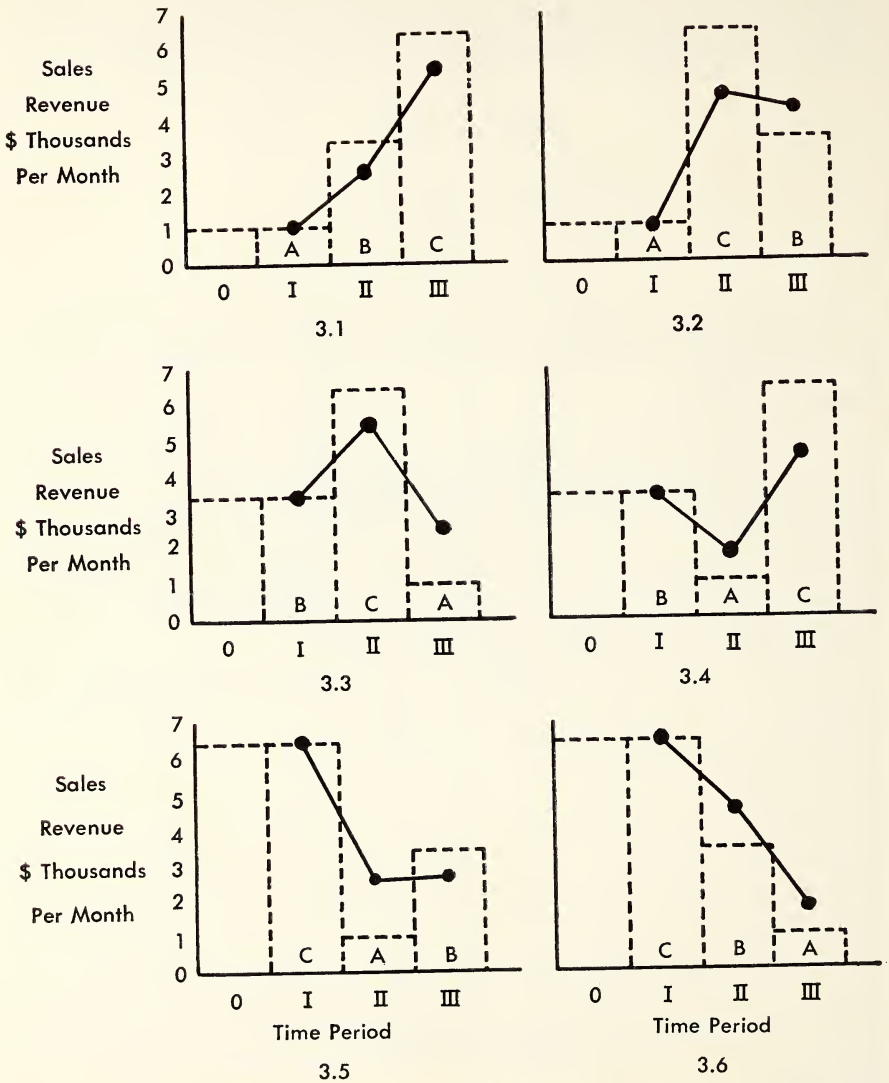


FIGURE 3. Illustrative response patterns. Broken line histograms indicate response assuming no carry-over. Small circles indicate observed response assuming carry-over.

great enough to make it worthwhile to complicate the statistical task that this design entails.

The general effect of lags is to obscure the underlying relationships. For example, in box 3.1 of Figure 3, the test area is presumed to have been run at the \$1,000 budget level (A) for the period before the experiment is regarded underway. The sales response in period I in both the

case of no-lag and lag-response, is the same—that is, \$1,000 per month. In period II, instant responses would provide sales of \$3,400 per month, but with lag, the observed response is \$2,680. Likewise, in period III the values are \$6,300 and \$5,430, respectively. In box 3.2 note that lag reduces the sales in period II, but increases it in III when the budget was reduced from the high to a medium level.

To look at the nature of response of sales to an advertising stimulus, let us consider a very simple model. Suppose we consider what happens if we put a small “dose” of advertising into a market. Following the advertising, we might expect that sales would rise to some peak and then trail off. We might represent this relationship by the simple graph at the top of Figure 4, where a dose of advertising was applied in the middle of period I

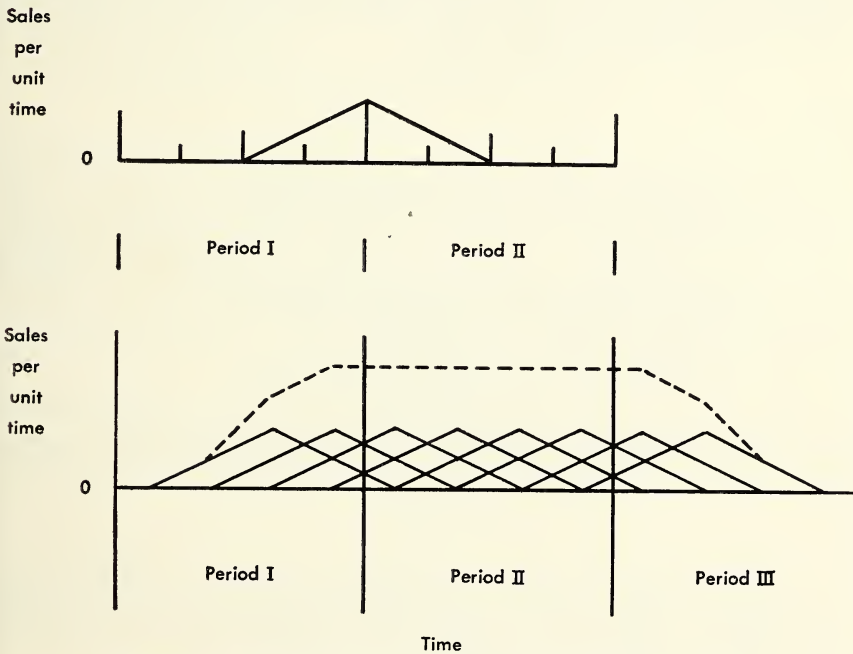


FIGURE 4. Simple model of sales response to advertising in doses.

and sales increased from zero to a maximum by the end of the period, two-fourths later, then trailed off to zero again in another two-fourths of a period. The total sales resulting from this dose of advertising is represented by the area under the curve (or the sum of the columns). This is the total response. However, only half the sales occurred in the period when the advertising was carried out. A “residual” of sales carried over to the subsequent period.

If we consider a series of doses during a given period and let each respond in this way independently of each other, we have a picture like

that at the bottom of Figure 4. Advertising outlays made during period I will build up to a maximum (in this case by the end of the period), and if the outlays are continued through period II, no further gain is made. Hence, in period II, a kind of equilibrium is reached where sales are at a maximum for this level of advertising outlay. If the advertising is withdrawn in period III, the sales will trail off to zero again. Later, we shall denote the sales in the first period as D , the direct response; those in the second period as T , the total response; and those in the third period as R , the residual response.

RESULTS

To provide illustrative material for Figure 3, a model was developed to represent the lag or carry-over of some advertising. Suppose we represent the amount of sales response observed in a given market, during a given period and at a given budget level, as a sum of a number of component effects:

$$y_{ijk} = D_k + R_k' + a_i + p_j + e_{ijk}$$

where

y_{ijk} = the observed response in the i^{th} market area during the j^{th} period when treatment k is being applied following treatment k' .

D_k = the actual response to treatment k during its period of application, or "direct" response.

R_k' = the actual response to treatment k' applied to the preceding period occurring during the present period, or "residual" response.

a_i = the differential effect of the i^{th} market area.

p_j = the differential effect of the j^{th} period.

e_{ijk} = all other effects, to be regarded as residual "error."

Since we are interested in the total sales effect of a given advertising application, we need to determine both D and R for a given treatment; that is, how much we get currently and how much subsequently. If we denote total response of treatment k by T_k , then

$$T_k = D_k + R_k$$

The switch-over type of design, with modifications, will permit unbiased estimates to be made of the total responses, T_k . As an illustration of the kind of data required and the computations involved, let us continue with the examples used so far. For the true values of T_k we have for treatment levels A, B and C:

$$T_A = 1.0 \text{ (thousands of dollars per month)}$$

$$T_B = 3.4 \text{ (thousands of dollars per month)}$$

$$T_C = 6.3 \text{ (thousands of dollars per month)}$$

Now, let us assume that direct and residual portions of the total responses to advertising are:

$$D_k = wT_k$$

$$R_k = (1 - w)T_k$$

where w is simply the fraction of the total response occurring during the period of the treatment, and $(1 - w)$ is the remaining fraction occurring in the subsequent period. If we assume for our example that $w = .70$ then,

Treatment

k	T_k	D_k	R_k
A	1.0	0.70	0.30
B	3.4	2.38	1.02
C	6.3	4.41	1.89

The observed responses for each of the six sequences of the three treatments is simply

$$y_k = D_k + R_k'$$

where R_k' is the residual from the previous treatment. The y_k 's for our example have been determined and plotted in Figure 3.

Let us assume that the period effects are:

$$p_1 = +0.1 \text{ (thousands of dollars per month)}$$

$$p_2 = 0.0 \text{ (thousands of dollars per month)}$$

$$p_3 = -0.1 \text{ (thousands of dollars per month)}$$

This means that during period I the sales response for \$1,000 per month of advertising outlay will be \$100 per month more than the average period; whereas, in period III, it will be \$100 per month less.

For market area effects we have (in thousands of dollars per month)

$$a_1 = -0.3 \quad a_4 = +0.1$$

$$a_2 = -0.2 \quad a_5 = +0.2$$

$$a_3 = -0.1 \quad a_6 = +0.3$$

That is, market area 1 during any period always responds with \$300 less sales per month for \$1,000 advertising outlay than the average market area; whereas market area 2 is \$200 less, etc.

Assuming residual errors are zero and aggregating the effects from the several sources, we have in Table 1 the values which would be observed in our switch-over experiment with the six test market areas. (Figure 3 is based on these data, except that area and period effects were omitted.)

TABLE 1
"OBSERVED" SALES IN THE SWITCH-OVER EXPERIMENT

<i>Period</i>	<i>Market Area</i>			<i>Total</i>
	<i>1</i>	<i>2</i>	<i>3</i>	
1	A = 0.80	B = 3.30	C = 6.30	10.40
2	B = 2.38	C = 5.23	A = 2.49	10.10
3	C = 5.03	A = 2.29	B = 2.48	9.80
	8.21	10.82	11.27	30.30
	<i>4</i>	<i>5</i>	<i>6</i>	<i>Total</i>
1	A = 1.20	B = 3.70	C = 6.70	11.60
2	C = 4.81	A = 1.92	B = 4.57	11.30
3	B = 4.27	C = 4.81	A = 1.92	11.00
	10.28	10.43	13.19	33.90

It may be of interest to see if the basic response data fed into the system can, in fact, be recovered. To do this, calculations of the following quantities are required:

H = total responses for all period-areas on which treatment k was put. This is the unadjusted total for each treatment.

F = total responses for all period-areas immediately following the one on which treatment k was put (including the pre-experiment treatment).

S_L = total of sequences where treatments occurred last.

S_F = total of sequences where treatments occurred first.

G = grand total of the experiment.

y = grand mean of the experiment.

The computed values of these quantities are shown in Table 2.

To obtain an estimate of T_k , the total sales attributable to the k^{th} advertising budget, both direct and residual, will be given by the estimator,¹

$$\hat{T}_k = \frac{H}{6} + \frac{1}{12}(3F - G + S_L - S_F)$$

For treatment A, the budget rate of \$1,000 per month, we obtain as an estimate of total sales.

$$\begin{aligned}\hat{T}_A &= \frac{10.62}{6} + \frac{1}{12}[(3)(16.48) - 64.20 + 24.01 - 18.49] \\ &= \frac{10.62}{6} - \frac{9.24}{12} \\ &= 1.00\end{aligned}$$

Hence, we have "recovered" the correct value of treatment A from our dummy experiment. In a similar manner the correct responses for treatments B and C can be estimated.

If we define the differential direct and residual effects in the system as

$$d_k = D_k - \bar{D} \text{ and } r_k = R_k - \bar{R}$$

respectively, where $\bar{D}$ and $\bar{R}$ are the means of the component effects respectively, then the true values for these quantities in this example are:

Treatment	d_k	r_k
A	-1.80	-0.77
B	-0.12	-0.05
C	+1.92	+0.82

and $\bar{D} = 2.50$, $\bar{R} = 1.07$ and $\bar{T} = 3.57$. To estimate r_k we have

$$\hat{r}_k = \frac{1}{12}[3F - G + S_L - S_F]$$

and

$$\hat{r}_A = \frac{1}{12}[(3)(16.48) - 64.20 + 24.01 - 18.49] = -0.77$$

To estimate d_k we have

$$T_k = \bar{T} + d_k + r_k$$

hence $\hat{d}_k = \hat{T}_k - y - \hat{r}_k$

$$\text{and } \hat{d}_A = 1.00 - 3.57 + 0.77 = -1.80$$

where $\hat{T}$ is estimated by y , the grand mean of the experiment. In general, we cannot estimate $\bar{D}$ and $\bar{R}$ individually with this design, but it is not of importance here anyway.

It should be noted that certain properties are peculiar to this design. The treatments are applied during the pre-experimental period so that observations made during the first period include residuals of known type. This also permits a simple method of estimating what is important here, the total responses T_k as well as the components, d_k and r_k . An estimate of w is also available from the experiment itself.

The variance of T_k is given by

¹The author is indebted to Dr. M. R. Mickey for this formula and for the appropriate method for computing "experimental error" for this design.

$$\hat{\text{Var}}(\hat{T}_k) = \frac{\sigma^2}{3}$$

where σ^2 is the appropriate experimental error for this design and

$$\hat{\text{Covar}}(\hat{T}_A, \hat{T}_b) = -\frac{\sigma^2}{12}$$

TABLE 2
COMPUTATION SUMMARY OF TOTAL, DIRECT, AND RESIDUAL EFFECTS

Treatment	<i>H</i>	<i>F</i>	<i>S_L</i>	<i>S_F</i>	$\hat{T}_k$	$\hat{d}_k$	$\hat{r}_k$
A.....	10.62	16.48	24.01	18.49	1.00	-1.80	-0.77
B.....	20.70	21.10	21.55	21.25	3.40	-0.12	-0.05
C.....	32.88	26.62	18.64	24.46	6.30	+1.92	+0.82
Totals.....	64.20	64.20	64.20	64.20	10.70	0.00	0.00
<i>G</i> = 64.20							
<i>y</i> = 64.20/18 = 3.567							

The appropriate method of computing the sums of squares for the analysis of variance table given in the previous section is conventional for all sources except residuals. Here the sum of squares is

$$4(\hat{r}_A^2 + \hat{r}_B^2 + \hat{r}_C^2)$$

The rather serious bias which can arise if the data are not treated properly, and which obscures the true relationship of sales to outlay, may be examined in this case. The unadjusted treatment totals are given by the T 's in the table below. From these we can compute the unadjusted treatment means. Thus, for the biased estimates of T_k we have,

$$T'_A = 10.62/6 = 1.77$$

$$T'_B = 20.70/6 = 3.45$$

$$T'_C = 32.88/6 = 5.48$$

When plotted against the true values (see Figure 5) the nature of the bias becomes apparent. In this case the response to low levels of outlay are overestimated and the high levels are underestimated. If guided by such results one may miss the true optimum level by a wide margin.

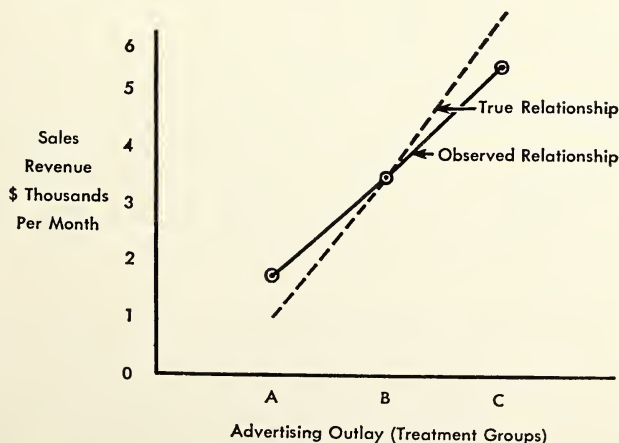


FIGURE 5. Distortion from carry-over effects.

DISCUSSION

The several steps still remaining will only be briefly mentioned here and perhaps dealt with in some detail in a later paper. The next step in this plan is to compute the Net Profit curve from the estimates of the portion of the Gross Sales curve covered by the experimentation (Figure 4), and from estimates of production costs. An examination of the information should be helpful to determine where the advertising budget should be set for all non-experimental market areas, and what levels should be chosen for further experimentation.

Note in Figure 6 that our illustrative experiment explored only a small portion of the Net Profit curve. Future experimentation would test

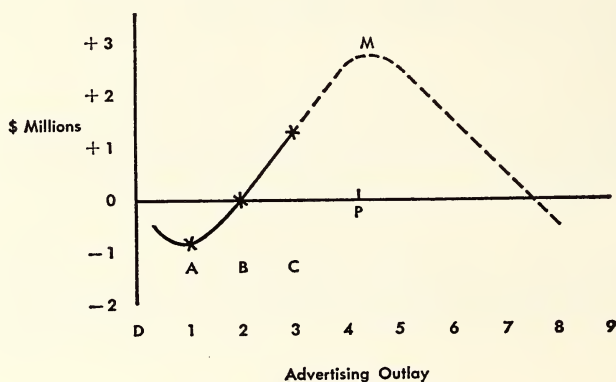


FIGURE 6. The net profit curve.

higher levels of advertising outlay in order to seek the maximum at M. (Bear in mind that the experimenter would not know the portion of the curve shown by a broken line.) He may be quite cautious, thinking that level C is near maximum already. The strategy to use at this point could be an interesting study for the future.

It is foolish to think that these relationships are fixed and once determined, will form the basis for all future action. Rather, it is more reasonable to regard the relationships as undergoing constant changes and that a business organization is faced with the problem of constantly seeking out the relationships *currently* in effect so that it can behave most rationally in regard to them. This requires that it keep up constant experimentation to determine where these curves are—perhaps, some day, to forecast them.

Some businesses will immediately see that a plan such as this one requires that sales be known by both market area and period. This will be a serious limitation unless measures are taken to obtain such information either from commercial services or by organizing their own. The advisability of either depends on expected gains against costs.

The basic ideas herein can be extended in many ways to deal with specific problems. For example, in a manner similar to that presented, an experimental plan could be made which would evaluate the effectiveness of different media rather than budget levels, or a combination of both. A more elaborate design could evaluate both media and budget levels. It appears that the response curves for different levels of inputs in the several media may differ somewhat, but what is probably more important are the interactions of media. Certain media mixes may have a total effect different from what the sum of the component parts might indicate. These differences can also be isolated and measured by appropriately designed experiments.

Sales Effects of Two Campaign Themes*

PETER L. HENDERSON, JAMES F. HIND,
and SIDNEY E. BROWN†

.....

SOUND PLANNING AND CONTROL OF ADVERTISING REQUIRES IMPROVED measurement of advertising effectiveness.

The purpose of this study was to evaluate the sales effectiveness of a specific promotional campaign. While the research techniques and statistical analyses were developed by physical scientists, the adaptation of these methods to advertising research has resulted in improved measurement of advertising effectiveness.

The basic experimental design and analysis was employed in the biological sciences as early as 1941 (Cochran *et al.*, 1941). It was later adapted and applied to problems in market research by Henderson (1952) as cited by Federer (1955) and in *Wood Chips* (1959).

A recent article published in the *Journal of Advertising Research* discussed the theory and concept of an experimental design as applied to a specific problem in advertising research (Jessen, 1961). A mathematical model of market simulation was described and computations made on simulated data. This paper reports the use of this design in a refined and extended form with real market data. It also discusses an entirely new technique for evaluating the effects of price, display allocation and customer traffic on the sales of advertised products.

The study was done in cooperation with the Washington State Apple Commission. During past years the Commission had developed a promotional program employing two advertising themes. One stressed the various uses of apples (fruit combination salads, baked apples, other dishes), while the other emphasized the healthful qualities of apples (builds strong bodies, dental benefits, etc.).

* Reprinted from the *Journal of Advertising Research*, Vol. I, No. 6 (December, 1961), pp. 2-11. Copyright 1961 by the Advertising Research Foundation, Inc., 3 East 54th Street, New York 22, New York.

† Agricultural Economists, U.S. Department of Agriculture.

The objectives of the study were to determine:

1. The over-all sales effectiveness of the promotional program relative to no promotion.
2. The relative sales effectiveness of the two promotional themes.
3. The short-time residual or carry-over effects, if any, of each theme.
4. The effects of the promotional activities for Washington apples on sales of apples from competing areas, and other selected fruits.
5. The influence on sales of apples and other fruits of various merchandising and advertising practices (*e.g.*, price cuts, display space and newspaper advertising) employed by stores to promote apples and other selected fruits.

METHOD

Experimental Design

Items tested, called treatments, included: an apple use advertising and promotional theme, a general health advertising and promotional theme and a control with no advertising or promotion. The control treatment was used to provide a basis for comparison. The experimental design to evaluate treatment effects was an extra-period double change-over design (see Table 1). Designs of this type are sometimes referred to as switch-back, switch-over or cross-over designs.

TABLE 1
EXTRA-PERIOD LATIN SQUARE CHANGE-OVER EXPERIMENTAL DESIGN USED IN
APPLE ADVERTISING STUDY OF 72 SUPERMARKETS IN SIX MIDWESTERN CITIES

Four-Week Time Periods	Sequence (Cities)					
	Square I			Square II		
	City 1	City 2	City 3	City 4	City 5	City 6
1.....	<i>A</i>	<i>B</i>	<i>C</i>	<i>A</i>	<i>B</i>	<i>C</i>
2.....	<i>B</i>	<i>C</i>	<i>A</i>	<i>C</i>	<i>A</i>	<i>B</i>
3.....	<i>C</i>	<i>A</i>	<i>B</i>	<i>B</i>	<i>C</i>	<i>A</i>
4.....	<i>C</i>	<i>A</i>	<i>B</i>	<i>B</i>	<i>C</i>	<i>A</i>

Treatment *A* = General health theme; Treatment *B* = Apple use theme; Treatment *C* = No advertising and promotion (control group).

The basic design (three time periods) consists of two replications of orthogonal Latin squares in which the sequence of treatments is reversed in the two squares. This design makes it possible to obtain estimates of direct and subsequent one-period carry-over effects of each treatment as well as the combined (direct plus carry-over) effects of each treatment with sustained use. The fourth time period, in which the treatments in the previous period are repeated, increases the precision of estimation of carry-over effects, which in turn results in greater accuracy for the estimates of the combined effects.

Treatments were assigned to successive four-week time periods

(rows) and cities (columns). Since each treatment occurred once in each row and column within a basic square, systematic errors resulting from constant variations among time periods and cities were equalized.

The mathematical model which forms the basis for the analysis is:

$$Y_{ijk t} = \bar{Y} + S_i + \underline{C}_{ij} + P_{ik} + D_t + R_{t(k-1)} + e_{ijk}$$

where:

- $Y_{ijk t}$ = Observed sales for the j^{th} city in the i^{th} square during the k^{th} period.
- $\bar{Y}$ = Over-all average sales of apples for all treatments and time periods.
- S_i = Effect of the square.
- $\underline{C}_{ij}$ = Effect of the city.
- P_{ik} = Effect of the period in that square.
- D_t = Direct effect of treatment.
- $R_{t(k-1)}$ = The carry-over (residual) effect of the immediately preceding treatment.
- e_{ijk} = Experimental error.

This is the analysis of variance model, and its underlying assumptions must be reasonably met for it to be properly employed in experimental research. These assumptions are reviewed elsewhere (Eisenhart, 1947). It is sufficient to note two basic assumptions: that constants in the model can be estimated without entanglement with each other, that is, they are additive; and that the experimental errors are independently and normally distributed.

Sample

Six midwestern cities, ranging in population from 100,000 to 150,000 and relatively free of sustained and intensive promotional campaigns for apples, were selected for the test. These cities were roughly comparable in major economic characteristics and supply considerations and overlapping of their local newspaper and television facilities was negligible.

Twelve self-service food stores were selected in each city to represent establishments of different sizes, different types of management and ownership (chains, voluntary chains and nonaffiliated independents), and different geographical areas of the city. Trade sources estimated that the panel stores in each city accounted for approximately 50 to 80 percent of retail food sales.

Intensity of promotion for the two promotional test themes (apple use and general health) featuring Washington State apples was as nearly equal as possible. Promotion included sponsored television programs Wednesday and Friday of each week and special tie-in advertising by retailers in media they normally used. During nonpromotional periods, retailers cooperated by following their normal merchandising and promotional practices for apples in the absence of a promotional campaign

by a commodity group. To insure reliable measurement of advertising themes, sample stores were asked to maintain: comparable apple displays for both test promotional themes; approximately equal store-sponsored promotion of apples by each test theme (special displays and features in newspaper ads); and comparable competition from selected fruits for each test theme in price, display area and feature advertising.

Weekly tonnage sales data for Washington State apples, apples from other areas, oranges, grapefruit and bananas were collected from each store by the standard audit method:

$$(\text{Beginning inventory} + \text{weekly receipts}) - (\text{ending inventory} + \text{transfers} + \text{withdrawals} + \text{spoilage}) = \text{sales.}$$

Additional data collected for each store on a weekly basis included total dollar sales in the produce department, and the amount of newspaper advertising by retailers for apples and selected fruits. Supplemental information on merchandising practices employed by the stores, such as prices and amount of display area for apples and other fruits, and the use of point-of-sale materials and special displays was collected by observation on Monday and Friday each week.

Analysis of Variance

Analysis of variance was used to separate the variations in sales of each fruit studied (Washington State apples, all apples, oranges, grapefruit and bananas) that were attributable to cities, time periods, direct effects, carry-over effects and experimental error. For illustration, the subdivision of the sum of squares for these attributes will be shown only for sales of all apples; the subdivision of sums of squares is similar for each fruit.

In the analysis of variance, only the partitioning of the sum of squares for direct and carry-over effects of treatments require special consideration. The conventional computations for the sums of squares for the other factors stratified in the experimental design and experimental error are given in textbooks on experimental design (Federer, 1955; Lucas, 1957).

Following the notation of Table 1, the sums of squares for the direct effects of treatments is given as:

$$\frac{1}{mn(n+1)(n+2)} \left\{ [(n+1)(\Sigma A) - (\bar{C}_2 + \bar{C}_6) - \text{G.T.}]^2 + \right. \\ \left. [(n+1)(\Sigma B) - (\bar{C}_3 + \bar{C}_4) - \text{G.T.}]^2 + \right. \\ \left. [(n+1)(\Sigma C) - (\bar{C}_1 + \bar{C}_5) - \text{G.T.}]^2 \right\}$$

with m being the number of squares (two), n the common number of rows, columns and treatments (three) in a square of the basic three time-period design, and G.T., the grand total of all observations.

Obviously, since no treatments involving promotion preceded those applied in the first time period, there are no carry-over effects of treatments in the first period. Thus, in computing the sum of squared deviations associated with carry-over effects, the numerical calculations should not include the data generated in the first period. The sum of squares for carry-over effects is computed as follows:

$$\begin{aligned} &\frac{1}{mn}[(B_{12} + A_{24} + B_{33} + C_{42} + C_{53} + A_{64})^2 + (C_{13} + C_{22} + B_{34} \\ &+ B_{44} + A_{52} + A_{63})^2 + (C_{14} + A_{23} + A_{32} + B_{43} + C_{54} + B_{62})^2] \\ &- \frac{1}{mn^2}(\text{G.T.} - \text{total period 1})^2 \end{aligned}$$

where the first subscript for each letter indicates the city and the second subscript indicates the time period.

The treatments designed in the first set of parentheses represent all treatments immediately following treatment *A* in cities one through six; similarly, treatments in the second parentheses represent those following treatment *B*, etc.

The complete partitioning of degrees of freedom and sums of squares for total apple sales (sales of apples from all areas combined) are given in Table 2. Carry-over effects in the analysis of total apple sales

TABLE 2
ANALYSIS OF VARIANCE, TOTAL APPLE SALES IN POUNDS

Source of Variation	Degrees of Freedom	Sums of Squares	Mean Square	F
Squares.....	1	2,417,070,246	2,417,070,246	55.83*
Periods within squares.....	6	2,437,012,149	406,168,692	9.38*
Between periods.....	(3)	(2,422,596,446)	(807,532,149)	18.65*
Periods x squares.....	(3)	(14,415,703)	(4,805,234)	0.11
Cities within squares.....	4	1,613,017,480	403,254,371	9.32*
Direct effect treatments....	2	492,948,044	246,474,022	5.69†
Carry-over effect treatments..	2	7,062,739	3,531,370	0.08
Error.....	8	346,325,670	43,290,709	
Total.....	23	7,313,436,328		

Under certain experimental conditions, the statistician might argue to pool the sums of squares and degrees of freedom for period x squares interaction (when insignificant) in the error term. However, when to pool and when not to pool is controversial and may tend to inflate the significance of treatment effects as in this case. Thus, for conservative estimates of treatment effects, this was not done.

* Significant at the 0.01 probability level.
† Significant at the 0.05 probability level.

did not approach statistical significance. Thus, the degrees of freedom and sums of squares associated with the carry-over effect were subsequently pooled in the error term as shown in a later table.

When significant carry-over effects are present, the following computational procedures are employed for estimating the direct and carry-over effect of each treatment.

Estimates of the direct effect per treatment, $\bar{D}_a$, $\bar{D}_b$, $\bar{D}_c$ are:

$$\bar{D}_a = \bar{Y} + \frac{1}{mn(n+2)}[n+1(\Sigma A) - (\mathcal{C}_2 + \mathcal{C}_6) - \text{G.T.}]$$

$$\bar{D}_b = \bar{Y} + \frac{1}{mn(n+2)}[n+1(\Sigma B) - (\mathcal{C}_3 + \mathcal{C}_4) - \text{G.T.}]$$

$$\bar{D}_c = \bar{Y} + \frac{1}{mn(n+2)}[n+1(\Sigma C) - (\mathcal{C}_1 + \mathcal{C}_5) - \text{G.T.}]$$

Estimates of carry-over effects per treatment, $\bar{R}_a$, $\bar{R}_b$, $\bar{R}_c$, are:

$$\bar{R}_a = \frac{1}{mn^2}[n(B_{12} + A_{24} + B_{33} + C_{42} + C_{53} + A_{64}) - (\text{G.T.} - \text{Total period 1})]$$

$$\bar{R}_b = \frac{1}{mn^2}[n(C_{13} + C_{22} + B_{34} + B_{44} + A_{52} + A_{63}) - (\text{G.T.} - \text{Total period 1})]$$

$$\bar{R}_c = \frac{1}{mn^2}[n(C_{14} + A_{23} + A_{32} + B_{43} + C_{54} + B_{62}) - (\text{G.T.} - \text{Total period 1})]$$

with symbols defined as before.

The estimate of the combined and direct carry-over of a particular treatment is the sum of the means of the direct and carry-over effects for that treatment, $\bar{D}_t + \bar{R}_t$, computed by the preceding formulae. Since advertising investment is usually spread over a sustained period of time, the combined direct and carry-over effects of treatments would appear to provide better estimates of treatment differences than direct effects alone (*i.e.*, when carry-over effects are significant).

For treatment contrasts, the standard error of a difference for direct, carry-over, and combined (direct plus carry-over) effects is given respectively as the square root of:

$$\frac{2(n+1)}{mn(n+2)} \cdot S^2, \frac{2}{mn} \cdot S^2, \text{ and } \frac{2(2n+3)}{mn(n+2)} \cdot S^2$$

where S^2 is the error mean square obtained in the analysis and m and n are as defined previously.

In the illustrative study, only the direct effects of treatments are considered, since carry-over effects were negligible. Estimates of direct effects of treatments, and the accompanying sample statistics, are given in Table 3 at this stage of the analyses (*i.e.*, before covariance analysis).

TABLE 3
TREATMENT MEANS AND SAMPLE STATISTICS
BEFORE AND AFTER COVARIANCE ANALYSIS*

<i>Item</i>	<i>Before Covariance Analysis</i>	<i>After Covariance Analysis</i>
	<i>Direct Effect ($\bar{D}_i$)</i>	<i>Direct Effect ($\bar{D}_i$)</i>
Treatment means:		
(Direct effects)		
Apple use theme.....	75,461	77,267
General health theme.....	71,575	69,722
No advertising or promotion.....	64,176	64,222
Sample statistics:		
Standard deviation†.....	5,945	5,333
Coefficient of variation (%)†.....	8.4	7.6
Standard error of a difference‡.....	3,070	2,754

* Treatment means are expressed as average sales per city per four-week time period.

† Based on pooled degrees of freedom and sums of squares for carry-over effects in the experimental error.

‡ Between two treatment means.

Analysis of Covariance

The sales data for Washington State apples and all apples were also adjusted for nonconstant sources of variation in sales that were not taken into account by the experimental design. These included variation in weather conditions among cities within time periods. The store's total dollar sales in the produce department were used in covariance analysis as an index of number of customers and purchasing power of customers in making adjustments. Produce sales did not completely satisfy the requirements of a concomitant observation, because produce sales and apple sales were not independent of each other since produce sales are affected to some extent by apple sales. However, apple sales contribute a relatively small percentage of the store's total produce sales. This was the best index for which data were available to reflect customer traffic and purchasing power.

The covariance correction for the regression of sales of all apples on total produce sales increased the precision of the findings because results were then based upon a constant number of customers and customer purchasing power in each city during each treatment. A brief description of the mathematical computations follows.

The analysis of sums of squares of produce sales, and the sums of products of produce sales and the sales of apples, is analogous to the analysis of variance previously described. The only difference is that in computing the cross products corresponding values of produce and apple sales are multiplied instead of squared in each stage of computation.

If Y_{ijkt} is defined as before and X_{ijkt} is the corresponding concomitant

observation for produce sales and b is the constant multiplier for the deviations of the concomitant variable from its over-all mean, the previous analysis of variance model takes the form:

$$Y_{ijkt} = \bar{Y} + S_i + C_{ij} + P_{ik} + D_t + D_{t(k-1)} + b(X_{ijkt} - \bar{X}) + e_{ijk},$$

with the symbols as defined before and $\bar{X}$ the over-all mean of produce sales.

The effects of the regression of produce sales (X) are removed from the sums of squares for error and direct effects treatment + error of apple sales (Y) by using the respective sums of squares for X and Y and corresponding cross products (XY) shown in Table 4 in the formula:

$$S_Y^2 - (S_{XY})^2/S_X^2$$

A degree of freedom is associated with regression and subtracted from degrees of freedom for error and treatment + error. The sum of squares of direct effect treatments, adjusted for produce sales, are then obtained by subtracting the corrected sum of squares for error from the corrected sum of squares for treatment + error. Degrees of freedom for treatments are similarly obtained. The complete computations can be followed in Table 4. A more detailed description of the standard computations involved are given in *Biometrics* (Cochran, 1957).

Adjusted treatment means are computed by: $\bar{D}_t - b(\bar{X}_t - \bar{X}_{ijk})$ where $\bar{D}_t$ is the mean apple sales of the t^{th} treatment, $\bar{X}_t$ is the corresponding mean for produce sales and $\bar{X}_{ijk}$ is the over-all mean of produce sales. The adjustment factor or regression coefficient b is given by the formula, $b = \frac{S_{XY}}{S_X^2}$ where S_{XY} and S_X^2 are the error sums of cross products (of X and Y) and the error sum of squares (for X) respectively, as shown in Table 4.

Adjusted treatment means, and other sample statistics computed before and after the covariance analysis are shown in Table 3. The increase in accuracy due to covariance analysis is demonstrated by the ten percent reduction in the size of the standard error of a difference. The use of the concomitant variable in the covariance analysis had almost the same effect as an additional Latin square in reducing the size of the difference between two treatment means which would be statistically significant. That is, at least one more replication of the four-period Latin square design would be required to attain the same precision in detecting significant differences if covariance were not used.

The F test was used to determine if differences in sales among the three treatment means were statistically significant. The least significant difference (LSD) test was used to determine significance between any two treatment means. These tests will not be discussed in detail, since

TABLE 4
ANALYSIS OF COVARIANCE, TOTAL APPLE SALES IN POUNDS (Y), PRODUCE SALES IN DOLLARS (X)

Source of Variation	Degrees of Freedom	Sum of Squares and Products		Error of Estimate*		
		S_X^2	S_{XY}	S_Y^2	Sum of Squares	Degrees of Freedom
Direct effects of treatments.....	2	36,021,777	42,438,284	492,948,044		
Error†.....	10	69,926,512	-82,535,043	353,388,409	255,971,376	9
Treatments + error.....	12	105,948,289	-40,096,759	846,336,453	831,161,596	11
Adjusted for produce sales.....					575,190,220	2
						287,595,110‡

$$* \text{ Sum of squares for error: } S_Y^2 - \frac{(S_{XY})^2}{S_X^2} = 353,388,409 - \frac{(-82,535,043)^2}{69,926,512} = 255,971,376$$

$$\text{Sum of squares for treatments + error: } S_Y^2 - \frac{(S_{XY})^2}{S_X^2} = 846,336,453 - \frac{(-40,096,759)^2}{105,948,289} = 831,161,596$$

† Carry-over effects are insignificant and are pooled in error terms.

‡ This treatment mean square with two degrees of freedom tested against error mean square with nine degrees of freedom is significant at the 0.005 percent probability level.

they are commonly used methods (Cochran, 1957). It should be noted, however, that a prerequisite to using the LSD test is the finding of significant differences among treatments by the F test.

Multiple Covariance Analysis

An extension of the covariance technique discussed previously was used to determine the nature and extent of the influence of retail merchandising practices and pricing policies on sales of apples (Washington State and other areas), oranges, grapefruit and bananas. Merchandising factors evaluated were price, display space and newspaper advertisement space. Produce sales, as before, were used to reflect number of customers and relative purchasing power of customers. These merchandising factors are referred to hereafter as quantitative factors. It was not possible to measure the direct effects on sales of such nonquantitative factors as variety, size and quality of fruits, size of pricing unit, type of display (prepackaged, bulk or combination) and packaging material.

Data for quantitative factors were tabulated for each city by weeks and plotted on scatter diagrams against sales of each fruit. Thus, a general indication was obtained of the factors related to volume of sales of each fruit. Factors which had no apparent relation to sales were eliminated.

Arithmetical computations in multiple covariance analysis are most laborious and complex. Ostle (1954) reviews them briefly. It would be quite boring and rather lengthy to discuss and illustrate the many iterative computations involved. These analyses are better explained by symbolizing a basic regression model and then following with some discussion of the adaptations and extensions of this basic model as used in the study.

The basic regression model is:

$$Y = \mu + b_1x_1 + b_2x_2 + \dots + b_nx_n + e$$

where Y is the sales observation in a city during any week for a particular fruit, μ the over-all mean and $b_1, b_2 \dots b_n$ regression coefficients, estimates of the effects of the $x_1, x_2 \dots x_n$ selected quantitative factors. The residual e is made up of city-time period and treatment effects, and random or experimental error. The regression model assumes that such quantitative factors as price, display area and newspaper advertisement space for the particular fruit remain constant between cities and four-week time periods. That is, none of the variation in sales is due to the influence of other variables associated with city and time period differences such as income levels, population characteristics and seasonal trends. This is highly unrealistic, since the influence of such variables vary considerably between cities and over time.

However, by combining the concepts of regression and analysis of variance, a more discriminating analysis can be made in which place and

season effects are removed. This technique, covariance analysis, permits the measurement of the net effects of these specified quantitative factors on sales. The analysis of variance model is:

$$Y_{ij} = \mu + P_i + C_j + e_{ij}$$

where Y_{ij} and μ are defined as before, P_i and C_j , the city and time period respectively. In this model the residual e_{ij} consists of the effects of the quantitative factors $b_1x_1 + b_2x_2 + \dots + b_nx_n$ and the random non-compensating errors of measurement. Combining the regression and analysis of variance models we have the covariance model:

$$Y_{ij} = \mu + P_i + C_j + b_1x_1 + b_2x_2 + \dots + b_nx_n + z_{ij}$$

where z_{ij} represents the random and noncompensating errors and the effects of the unidentified factors. This model, unlike the previously stated regression model, defines and accounts for the effects of city and seasonal differences. Thus, the estimates of the effects $b_1, b_2 \dots b_n$ of the quantitative factors $x_1, x_2, \dots x_n$ on sales are free of the place and season effects.

Based on the covariance model, a multiple analysis of covariance was first used to adjust the sales variation for each fruit for the variations associated with cities and time periods. A multiple regression analysis was then made of the adjusted data (i.e., the residual sums of squares and cross products) to identify and quantify the net effects of the merchandising factors significantly affecting sales. The multiple regression analysis was repeated until only those factors affecting sales remained which attained statistical significance at the 0.05 probability level. The practical utility of a factor based upon the magnitude of its regression coefficient (b value) and coefficient of determination (R^2) was also a criterion for retaining it in subsequent analysis. The complete analysis for Washington State apple sales is shown in Table 5.

RESULTS

There were substantial differences in sales of both Washington State and all apples between periods with promotional themes (apple use and health) and periods of no promotion (see Table 6).

When sales of apples were combined, the apple use theme was significantly more effective in promoting sales than the health theme. However, the nine per cent sales difference between the two themes for Washington State apples was not large enough to be statistically significant.

The themes used in the four-week test period significantly affected neither Washington State *nor* total apple sales during the next four-week period without advertising.

TABLE 5
MULTIPLE COVARIANCE ANALYSIS

<i>Source of Variation</i>	<i>Degrees of Freedom</i>	<i>Sums of Squares</i>	<i>Mean Squares</i>	<i>F*</i>
Total.....	863	412,284,650		...
Nonquantitative factors:				
Cities and time periods.....	17	129,383,870	7,610,816	44.8
Residual 1.....	846	282,900,780	334,398	
Quantitative factors:†				
Display space for Washington State apples.....	1	48,705,457	48,705,457	286.6
Produce sales‡.....	1	41,561,786	41,561,786	244.5
Display space for other apples.....	1	4,861,877	4,861,877	28.6
Price of Washington State apples.....	1	3,630,474	3,630,474	21.4
Display space for grapefruit.....	1	2,558,838	2,558,838	15.1
Newspaper advertisement space for Washington State apples.....	1	2,113,867	2,113,867	12.4
Joint effects of above factors (interactions).....	4	36,697,381		
Residual 2.....	840	142,771,100	169,996	

* All significant at the 0.01 percent probability level.

† Quantitative factors adjusted for effects of each other and nonquantitative factors of cities and time periods.

‡ Used as an index to reflect changes in the number of customers patronizing the store and changes in the relative purchasing power of customers.

Advertising Washington State apples exerted only a minor influence on sales of oranges, grapefruit and bananas. The effect of the advertising seemed to vary among the fruits depending on the theme employed. The sales differences were too small to determine whether the promotional themes for apples significantly improved the sales of these fruits or not. The differences that were found corresponded to findings of previous research studies, namely, that advertising and merchandising practices which increase sales of apples also benefit sales of oranges, as suggested by the data in Table 7. Also, the decrease in sales of bananas when either apple theme was advertised compared to no promotion was similar to findings of previous studies which have indicated that apples and bananas are competitive products (Henderson, 1952, 1955, 1955).

Changes in sales of apples, oranges, grapefruit, and bananas were significantly related to changes in some but not all of the practices employed by stores in merchandising and promoting these fruits, such as amount of display area, newspaper advertising, and prices (see Table 8).

Sales of each fruit were generally affected by the merchandising and promotional practices used with it. The major exception was the amount of display space devoted to each fruit, which affected the fruit displayed and also had some influence on other fruits. Variation in the amount of display space used for grapefruit affected all fruit except bananas. Grapefruit sales varied directly with the amount of space in grapefruit displays, while sales of apples and oranges varied inversely with the

TABLE 6
SALES OF APPLES FROM WASHINGTON STATE AND OTHER AREAS DURING NO APPLE PROMOTION AND TWO APPLE PROMOTIONAL THEMES¹

Source of Apples	Average Sales per Store per Four-Week Period				Difference in Sales between No Promotion and— ²					
	With No Promotion		With Apple-Use Theme		With Health Theme		Apple-Use Theme and Health Theme			
	Lb.	Lb.	Lb.	Lb.	Lb.	Lb.	Lb.	Lb.	Lb.	%
							%		%	
Washington State.....	3,124	4,117	3,784	993†	31.8	660*	21.2	333	8.8	
Other areas (Midwestern and Eastern states and Canada).....	2,227	2,322	2,026	95	4.3	-201	-9.1	295	14.6	
Total or average.....	5,351	6,439	5,810	1,088†	20.1	459*	8.6	629†	10.8	

¹ All sales data were adjusted for variations among treatments (themes and no promotions) which might be attributed to differences in number of customers and purchasing power per customer. While there were price reductions associated with weekly specials featured by retailers, average prices (weighted by sales) did not differ significantly among treatments.

² Differences required for statistical significance for apples from Washington State and from other areas are ± 735 pounds at the 5 percent probability level; ± 596 pounds at the 10 percent probability level.* Differences required for statistical significance for apples from all areas combined are ± 748 pounds at the 1 percent probability level; ± 520 pounds at the 5 percent probability level; and ± 422 pounds at the 10 percent probability level.*

The economic implication of these probability levels is that the chances or odds of obtaining sales differences of these magnitudes are less than 1, 5, or 10 in 100 from random sales fluctuations and nonconstant experimental errors.

TABLE 7

SALES OF SELECTED FRUITS DURING NO APPLE PROMOTION AND TWO APPLE PROMOTIONS*

Fruit	Average Sales per Store per Four-Week Period			Difference in Sales between No Promotion and—†	
	With No Promotion	With Apple Use Theme	With Health Theme	Apple Use Theme	Health Theme
	Lb.	Lb.	Lb.	%	%
Oranges.....	5,516	5,680	5,784	3.0	4.9
Grapefruit.....	5,272	5,156	5,996	-2.2	13.7
Bananas.....	5,944	5,836	5,752	-1.8	-3.2
Total.....	16,732	16,672	17,532	-0.4	4.0

* All sales data were adjusted for variations among advertising treatments which might be attributed to differences in number of customers and purchasing power per customer. Sales data were further adjusted for the effects of prices, display space, and other significant merchandising and promotional practices employed by stores. The regression coefficients obtained in the multiple regression analysis were used in adjusting means. Computations are similar to those made in obtaining the adjusted means in the covariance analysis.

† Differences required for statistical significance at the 5 percent probability level are ± 10.8 percent for oranges; ± 17.5 percent for grapefruit; and ± 10.0 percent for bananas. Differences required for statistical significance at the 10 percent probability level are ± 8.7 percent for oranges; ± 14.2 percent for grapefruit; and ± 8.1 percent for bananas.

TABLE 8

MERCHANDISING AND PROMOTIONAL PRACTICES WHICH SIGNIFICANTLY AFFECTED SALES OF WASHINGTON STATE APPLES, APPLES FROM OTHER AREAS, AND OTHER SELECTED FRUITS

Factors	Factors Significantly Affecting Sales of—*				
	Washington State Apples	Other Apples†	Oranges	Grapefruit	Bananas
Produce sales (Dollars).....	+ 0.3	+ 0.3	+ 0.7	+ 0.8	+ 0.7
Price (Cents per pound):					
Washington State apples....	- 25.9	...	...	...	...
Other apples.....	...	- 8.8	...	...	...
Oranges.....	...	...	- 93.9	...	...
Grapefruit.....	...	...	...	- 122.1	...
Bananas.....	...	...	...	...	- 86.6
Display space (Square feet):					
Washington State apples....	+ 21.8	- 8.4	...	...	...
Other apples.....	- 6.8	+ 25.8	...	...	...
Oranges.....	...	...	+ 16.2	...	...
Grapefruit.....	- 4.8	- 6.4	- 7.3	+ 10.7	...
Bananas.....	...	...	- 14.9	...	+ 15.6
Newspaper advertising space (Square inches):					
Washington State apples....	+ 5.0	...	...	...	...
Oranges.....	...	...	+ 6.5	+ 8.5	...
Grapefruit.....	...	...	...	...	...
Bananas.....	...	...	...	...	+ 19.6

* A plus sign indicates that a positive change (increase) in the value of a factor is accompanied by an increase in sales of a fruit and a negative change by decrease in sales. For example, on the average, an increase of one square foot in the display space for Washington State apples was accompanied by an increase of 21.8 pounds in Washington apple sales. A negative sign signifies that a positive change in the factor results in decreases in Washington apple sales and a negative change (decrease) in the factor results in an increase in Washington apple sales. Thus, on the average an increase of one cent a pound in the price of Washington apples resulted in a decrease of 25.9 pounds in sales of Washington apples.

† Apples from areas other than Washington State.

amount of space for grapefruit displays. Banana sales increased and orange sales decreased when the size of banana displays was increased. Similarly, a decrease in banana sales and an increase in orange sales were associated with a decrease in the amount of space devoted to banana displays. These findings indicate that grapefruit competes with apples and oranges for display space and sales, while bananas compete with oranges for display space and sales.

Price and display space devoted to each fruit exerted the most influence on sales. Sales for each fruit varied directly with the amount of display space and inversely with price. The variation in sales of each fruit from week-to-week was also generally related to the week-to-week variation in amount of newspaper advertising space devoted to each fruit by retailers.

Sales of each fruit were significantly and directly related to the volume of produce sales. Produce sales reflect the combined effects of promotional and merchandising practices employed by stores on the sales of individual products, and the influence of such practices in drawing additional customers into the stores.

OTHER APPLICATIONS

While the notation and terminology used refer to this specific experiment, the design and analyses presented would be valid in other areas of advertising research where measurable carry-over effects of promotional techniques are likely to occur. Some general applications are given in the following illustrations.

This experimental design could be used to predict the most efficient promotional alternatives from consumer or trade media advertising, personal selling, point-of-purchase effort, cooperative advertising and premium offers, and to determine the place of each in the total promotional effort for a product. At the same time, holding promotional expenditures relatively constant for each technique tested, promotional outlay could be related to sales returns: Using a more complicated arrangement of this design, the optimum levels of promotional outlay under different conditions could also be found.

With these alternative uses of the design the multiple covariance technique could be used to obtain estimates of sales responses (for a product sold at retail) to changes in prices, display space and other merchandising practices employed by retail stores. This information could be used in deciding what combination of price and merchandising practices will tend to maximize returns from advertising and promotional activities. The covariance technique could also determine relationships between sales of a product and such factors as price of the product, price of related or competing products, and consumers' incomes, in order to forecast sales in a market from measurements on a sample. The estimated parameters for such quantitative variables could be applied to other

market areas with similar socio-economic characteristics, since these estimates would be corrected for place and season effects.

REFERENCES

- COCHRAN, W. G., *et al.* "A Double Change-over Design for Diary Cattle Feeding Experiment," *Journal of Dairy Science*, Vol. XXIV, No. 11 (November, 1941), pp. 937-51.
- COCHRAN, W. G. "Analysis of Covariance: Its Nature and Uses," *Biometrics*, Vol. XIII, No. 3 (September, 1957), pp. 268-69.
- COCHRAN, W. G. *Experimental Designs*. 2d ed. New York: John Wiley & Sons, Inc., 1957, p. 76.
- EISENHART, CHURCHILL. "The Assumptions Underlying the Analysis of Variance," *Biometrics*, Vol. III, No. 1 (March, 1947), pp. 1-21.
- FEDERER, WALTER T. *Experimental Design*. New York: The Macmillan Co., 1955, pp. 148-49 and 449-52.
- HENDERSON, PETER L. "Application of the Double Change-over Design to Measure Carryover Effects of Treatments in Controlled Experiments," *Methods of Research in Marketing, Paper No. 3*. Ithaca, N.Y.: Cornell University, 1952.
- HENDERSON, PETER L. "Marketing Services Affecting Apple Sales." (Unpublished Ph.D. dissertation; Ithaca, N.Y.: Cornell University Library, 1952).
- HENDERSON, PETER L. "Merchandising Apples," *Virginia Fruit*. Virginia Horticultural Society, Vol. XLII (May, 1955), pp. 34-36.
- HENDERSON, PETER L. "The Effect of Advertising and Point of Sale Aids on Apples," *Virginia Fruit*, Vol. XLII (June, 1955), pp. 30-34.
- JESSEN, RAYMOND J. "A Switch-over Experimental Design to Measure Advertising Effect," *Journal of Advertising Research*, Vol. I, No. 3 (March, 1961), pp. 15-22.
- LUCAS, H. L. "Extra-period Latin Square Change-over Design," *Journal of Dairy Science*, Vol. XL, No. 3 (March, 1957), pp. 225-39.
- OSTLE, BERNARD. *Statistics in Research*. Ames, Iowa: The Iowa State College Press, 1954, pp. 406-8.
- WOOD RESEARCH CORPORATION. "The Experimental Method in Marketing Research," *Wood Chips*, February, 1959.

A Technique for Studying the Effects of a Television Broadcast*

.....

WILLIAM A. BELSON†

Introduction

IN THE SUMMER OF 1953 THE BBC'S TELEVISION SERVICE BROADCAST A SERIES of four programmes called "Bon Voyage," the aims of the series being (i) to teach viewers French words and phrases of a kind likely to be of use to them in making a first trip to France; (ii) to provide them with general information for the same purpose; (iii) to remove apprehensions about possible language difficulties. An intensive study was made of the extent to which these aims were achieved, and the general findings are reported elsewhere.¹ This paper, however, is concerned solely with the *research technique* used in that inquiry. Whereas the technique was used specifically to assess the effects of the "Bon Voyage" series, and is described in that context, it appears to have value as a general research tool for the study of effects.

A research design often advocated and used with advantage in studies of effects requires that people be tested before and after exposure to the influencing conditions.⁵ The assessment of the effects of a television broadcast is, however, beset by difficulties which preclude the use of a simple *before-and-after* design. In the first place, it is essential that the viewing upon which the results are to be based should be entirely normal (i.e. not influenced either by pre-broadcast testing or by any knowledge that tests are subsequently to occur). Secondly, it is often impossible with a television broadcast to say in advance of the programme precisely what its content will be—as would be necessary for any effective *before-and-after* comparison.

Of course, a superficial way out of these two difficulties would have been to conduct *all* the test work *after* the series had been broadcast,

* Reprinted from *Applied Statistics*, Vol. V, No. 3 (November, 1956), pp. 195–202.

† The London School of Economics and Political Science.

¹ See reference 1 at end of article.

the method being to compare the test scores of those people who *did* see the programme with those of non-viewers or persons who did *not* see it. The difficulty here, however, was that the difference between the two groups could not have been interpreted as pure effect, but only as a mixture of effect and of original (or pre-broadcast) difference between the two groups of people. After all, one of the groups elected to view the programme while the other did not, and (apart from those who might have wished to view but could not) this presupposes some difference in attitude towards the subject of the programme. Accordingly, for the *after-only* design to be of value, methods had to be found for partialling out or eliminating all or most of that part of the difference between test scores which was attributable to the fact that the groups were different to start with. Such a method was in fact devised and was used with considerable success in the "Bon Voyage" study. (As a method it had its roots in a form of partial correlation which I had used earlier for reducing volunteer bias in invited groups and for general matching purposes.² Wilkins⁶ has developed techniques of a related kind in the field of criminological prediction.)

The Rationale of the Method

Basically, the method used was an adaptation of the more usual matching technique, though unlike ordinary matching methods (i) there was no discarding of subjects, (ii) the matching criteria were established empirically, (iii) it was possible to make a relatively firm estimate of the success of the method in eliminating extraneous differences.

The rationale of the method may be summed up in the following argument. Suppose that the non-viewers' knowledge of the broadcast French words is predictable (with a known degree of accuracy) through a particular set of variables *which are not themselves open to influence by the programme*. Suppose also that the viewers and the non-viewers differ in respect of these prediction variables. Then, using a regression equation based upon the prediction variables, it is possible to estimate what the non-viewers' score on the broadcast French words *would* have been had they equalled the viewers in respect of the prediction variables. The residual difference between the scores of the two groups then approximates to "effects" of the programme. Clearly the higher the multiple correlation between the prediction variables and the variable being studied (e.g. knowledge of the broadcast words), the greater is the confidence that the residual difference is "pure effect" (and not partly a result of extraneous differences). Accordingly the ideal in using the technique must be to build up the multiple correlation to as high a level as possible. Strictly speaking, a multiple correlation of anything less than ± 1.0 leaves room for doubt about the interpretation of residual differences, for it is still possible that the remaining variance springs from the operation of variables in respect of which the viewers and the non-viewers differ. It

is therefore most important in using this technique to incorporate into the design a means of making a check upon the degree to which the chosen prediction variables have done their job.

The use of this approach in the "Bon Voyage" study thus required the development of both high-powered predictors and an effective checking device; some details of these are set out in the next two sections. Since it is a vital feature of the technique that the prediction or matching variables should not themselves be open to influence by the programme, they are referred to hereafter as the "stable correlates" of the variable under study, and the method itself may for convenience be called the "stable correlate" method.

Developing the Stable Correlates

As is typical of the method, no additional test-sessions were required for the development of the stable correlates. It was necessary in the first place to postulate a number of variables which it was reasonable to think were associated with (and therefore predictive of) the variable under study. For example, "number of trips made to France" was expected to be predictive to some extent of knowledge of the French words broadcast. Measures in terms of some six of these proposed correlates were then made *during the post-broadcast test sessions* (along with the main tests). After this, a preliminary analysis was made to determine, statistically, precisely what combination of the proposed stable correlates was most predictive of test score. Since in this particular project three separate variables were under study, three separate analyses had of course to be made, the possibility being that three different combinations of correlates would be needed. It was essential to restrict these analyses to the test results of the non-viewers because the inclusion of those of the other group would, if the programme had in fact produced changes, have led to a misleading attenuation of the correlations. The calculation was made through the Wherry-Doolittle formula⁴ and, as is usually the case, two to three of the proposed stable correlates taken in combination came very close to maximising correlation. Thus in Table 1, which sets out the main results of this preliminary analysis, the two main predictors of "knowledge of the broadcast French words" yielded a multiple correlation of +0.82, and this figure would have been raised only slightly (+0.85) by the addition of a third stable correlate.

This made it possible to develop, for each of the three variables under study, a short regression equation which, as already stated, was to be used to adjust the score of the non-viewers. Such a correction formula, based upon two variables x_1 and x_2 , would take the form

$$\bar{y} = \beta_1 x_1 + \beta_2 x_2$$

(where $\bar{y}$ is the corrected score and β_1, β_2 are constants, all the variables being measured from their means), although, as in the present case,

TABLE 1
THE PROPOSED MATCHING CRITERIA

<i>Variable under Study</i>	<i>Stable Correlates Selected (in Order of Priority)</i>		<i>Multiple Correlation Achieved</i>
	<i>Variable</i>	<i>Correlation</i>	
Knowledge of French words and phrases presented in "Bon Voyage"	Control words	+0.78	+0.82 (first two)
	Educational level	+0.59	+0.85 (all three)
	Occupational level	+0.58	
Knowledge of facts presented in "Bon Voyage"	Control words	+0.54	+0.63 (first two)
	Visit to France*	+0.46	+0.65 (first three)
	Educational level	+0.51	+0.65 (all four)
	Occupational level	+0.43	
Attitude on issues related to visiting France	Control words	+0.48	+0.52 (first two)
	Visit to France	+0.34	+0.55 (all three)†
	Occupational level	+0.39	

* Those who had made a visit in the past were more favourably and confidently disposed to making a further visit.

† To have selected more correlates than those shown would, in the case of each of the variables under study, have involved no appreciable increase in the multiple association (see Wherry-Doolittle formula).

whenever the various distributions are not strictly normal or the associations not strictly rectilinear, a simple weighting system as described below is preferable. Take for example the case of "knowledge of the broadcast French words," where the matching criteria are "control words" and "educational level." If each of these variables had allowed a division of subjects into three groups, then it would have been possible to classify the 120 subjects of the non-viewing group into nine sub-groups (e.g. control words X, educational level A; control words Y, educational level B; and so on), and these nine groups would then have been the relevant ones as far as matching was concerned. The aim of the weighting process would then be to determine what the score of the total non-viewing group *would* have been if each of its nine sub-groups had contained the same number of people as were in the comparable sub-groups of viewers.

The effect of this correction, made through the weighting method, is illustrated in Table 2. Thus the unadjusted average score of non-viewers on the word knowledge test was 4.89 out of the total of 26; on adjustment through the appropriate correlates this became 5.31, which then compared with the average score of 6.21 by the viewers—indicating that the programme had produced an increase (significant at the 0.02 level) in knowledge of the broadcast words. To illustrate further, the adjustment to the non-viewers' attitude score (through *its* correlates) moved that average from 15.38 to the more "favourable" average of 17.65, which then compared with the viewers' average of 11.80—indicating that the programme had produced a sharp decrease in viewers' confidence about going to France (significant at the 0.01 level). From this and the re-

TABLE 2
CHANGES PRODUCED (POST-BROADCAST GROUPS)

<i>Variable under Study</i>	<i>Nonviewers' Scores</i>		<i>Viewers' Scores</i>	<i>P</i>
	<i>Unadjusted Averages</i>	<i>Adjusted Averages</i>	<i>(Unadjusted Averages)</i>	
Knowledge of words presented in "Bon Voyage" (creative response)*.....	4.89	5.31	6.21	0.02
Knowledge of facts presented in "Bon Voyage"†.....	8.54	9.08	10.10	0.02
Attitude on issues related to visiting France‡.....	15.38	17.65	11.80	0.01

* Score out of 26.

† Score out of 29.

‡ 17.65 represents a relatively favorable and confident attitude.

Note. These figures are unadjusted for "attendance bias" (that is, less than half of the number invited actually attended); where such adjustments *are* made, however, the estimates of amount of change produced are virtually unaffected.

maining data in Table 2 it will also be seen that in two of the adjustments the result was a reduction in the apparent gap between viewers and non-viewers, while in the third it was to increase the gap.

Assessing the Adequacy of the Correction

Under favourable conditions a relatively high correlation can be developed when the variable being studied is dependent upon an ability of some kind, though it may, as instanced in Table 1, be as low as +0.5 or +0.6 in the case of an attitude (e.g. attitude towards going to France). A multiple correlation as low as this may well be a reflection of multi-dimensionality in the "variable" being studied rather than a failure to identify relevant matching variables; but in any case, however high the multiple correlation, a check on the adequacy of any matching combination is essential.

It is typical of the method that this check is indirect, and this was the case in the "Bon Voyage" study. Had it been feasible—as obviously it was not—to test the two groups before the broadcast and then to apply the correction formula to the score of the non-viewing group, an accurate estimate of the adequacy of the correction could have been made, for obviously a fully effective correction formula would have removed *all* pre-broadcast difference. Something approximating to this type of check *was* feasible, however, and was made without interfering with the research design. It was necessary to make the pre-broadcast tests on two groups of people approximating to the two post-broadcast groups as closely as expectations allowed. Full details of the methods used to secure these groups are given later, and for the present it will suffice to say that, like the people taking part in the post-broadcast tests, their names were available from the records collected in the department's Daily Survey of

Listening and Viewing. *It must be clearly pointed out, however, that once these pre-broadcast groups were tested they did not take part in any way in further tests.* Their one and only function was to provide a check on the adequacy of the correction formulae, and the post-broadcast groups were made up of entirely different people. It must also be noted that since the programme's content could not, in advance of the broadcast, be known exactly, the pre-broadcast tests could be only approximations to the post-broadcast tests, though there was nothing to prevent them from being very close approximations. The check through the pre-broadcast groups could only be applied, of course, after the correction formula had been developed through one of the post-broadcast groups. The necessary adjustments to pre-broadcast scores were then made and any residual differences noted. Results are set out in Table 3, and from

TABLE 3

ADEQUACY OF THE STABLE CORRELATES

(Showing Matching Corrections Applied to the Scores of Two Groups, Viewers and Nonviewers, Tested before the Broadcast of "Bon Voyage")

<i>Variable under Study</i>	<i>Nonviewers' Scores</i>		<i>Viewers' Scores</i>
	<i>Unadjusted Averages</i>	<i>Adjusted Averages</i>	<i>(Unadjusted Averages)</i>
Knowledge of French words and phrases . . .	4.43	4.26	4.26
Attitude on issues related to visiting France	6.77	6.66	6.60

these it will be seen that even in the case of the attitude assessment the result of the correction was to eliminate pre-broadcast differences almost completely. A high degree of confidence could therefore be entertained that the residual differences between the two post-broadcast groups (see Table 2) did in fact represent "effects."

The Securing of Subjects

Although a description of the method used to secure subjects is not essential to the exposition of research design, it is necessary for a balanced description of this particular application of it. The securing of subjects was, without a doubt, simplified and made inexpensive by the special facilities of Audience Research. At the same time these facilities are by no means essential and are replaceable by preliminary survey work.

a) *Post-broadcast Groups.* The department's daily survey of listening and viewing provides the names and addresses of a sample of people who heard or saw programmes broadcast on the day preceding the survey, and records also each person's age, sex and social group. This made it possible to identify two groups for post-broadcast testing: (i) a group which had seen the programme; (ii) a group of non-viewers who

had not seen the programme and who were approximately the same as the other group in terms of age, sex and social group.

b) Pre-broadcast Groups. "Bon Voyage" occurred as a section of a regular afternoon programme called "Leisure and Pleasure," so that the names and addresses of potential viewers of "Bon Voyage" were available for invitation to viewer meetings before the "Bon Voyage" broadcast. A group of non-viewers, similar in terms of age, sex and social group, could also be recruited (for pre-testing) from the survey records.

Some Other Applications of the Method

Although I developed this particular "stable correlate" adaptation for the study of the effects of specific programmes, I have since used it in a study, using survey methods, of some of the long-term psychological effects of television,³ and it appears to be readily applicable to other sociological enquiries. Indeed, I should like to suggest this adaptation as a general research technique for use in those studies of "effects" where circumstances preclude the making of measurements "before the event." Such circumstances are only too common and occur: (a) where pre-testing is likely to affect subsequent exposure to the influencing conditions and the performance on the final test; (b) where pre-testing is in any case out of the question because either (i) there is no knowing "before the event" precisely who will be exposed or elect to be exposed to the influencing conditions, or (ii) the effect to be studied has either occurred or begun to occur before the commencement of the study (as in sociological enquiries into the effects of cinema-going, publicity campaigns, and so on). Such circumstances do not necessarily enforce an *after-only* design, but where they do, the limitations of that design *as normally applied* are generally well recognised. The very least that the proposed adaptation can do in such studies is (i) to remove some of that part of the difference (between groups) which is not an effect of exposure to the events under study and (ii) to provide some indication of the degree to which that extraneous difference has in fact been removed. This in itself seems to me to be a useful advance.

It is tempting, however, in the light of the "Bon Voyage" study, to go further than this. At its best—that is, granted high-powered correlates and an efficient checking device—the adaptation could make possible a relatively unambiguous assessment of effects. A realistic approximation to this condition can never be guaranteed, of course, because in the end the origination of good correlates is dependent upon the individual research worker and upon the nature of the variable being studied. At the same time it is usually possible in sociological enquiries to spend a little time beforehand in developing, testing, selecting, and improving the proposed correlates, and this can contribute greatly to their final predictive power. Another difficulty is that the less controlled are the nature of the influencing events and the conditions of exposure to them, the less

direct is likely to be the available device for checking the adequacy of the corrections made—so that up to three or four independent checks may in fact have to be devised.

Granted all this, however (and the demands are not excessive), the proposed adaptation can make of an *after-only* design a means of achieving reasonably accurate assessments of effects without interfering in any way with that essentially multi-determinant context which is likely to surround and to condition the normal operation of the factor under study.

REFERENCES

1. BELSON, W. A. "Learning and Attitude Changes Resulting from Viewing a Television Series 'Bon Voyage,'" *British Journal of Educational Psychology*, Vol. XXVI (1956), p. 31.
2. BELSON, W. A. *The Comprehensibility of "Topic for Tonight."* Audience Research Department (BBC) Report, LR/52/1080, 1952.
3. BELSON, W. A. "The Effects of Television upon the Interests and Initiative of Viewers" (study nearing completion), 1956.
4. GARRETT, H. E. *Statistics in Psychology and Education*. New York: Longmans, Green & Co., Inc., 1947.
5. HOVLAND, C. I., *et al.* *Experiments on Mass Communication*. Princeton, N.J.: Princeton University Press, 1949.
6. WILKINS, L. T. "Some Practical Applications of Recent Developments in Statistical Theory," *The Incorporated Statistician*, Vol. IV (1953), p. 92.



**READINGS ON
STATISTICAL
ANALYSIS**

PART

FOUR



Cross- Classification Analysis

THE TECHNIQUE OF CROSS-CLASSIFICATION is designed to organize data into a form that allows fast and easy interpretation. Cross-classifications take the form of numerical tables or of charts and graphs, from which the desired relationships can be inferred. While the technique is particularly useful in exploratory studies, where the problem is not defined well enough to permit more formal methods to be applied, cross-classifications are valuable adjuncts to almost any statistical analysis. Examples of two different approaches to cross-classification are presented in the readings.

The first article was prepared by the editors of *Wood Chips*, a publication of the A. J. Wood Research Corporation. Entitled "Research Design," it was written from the point of view of the practicing market research analyst in business. In it, cross-classification is viewed as the central technique for inferring causal relationships from a set of market data. The fact that the article was prepared to illustrate the use of cross-classification serves to enhance its interest.

Two basic principles must be considered when running cross tabulations on a given set of data. The editors of *Wood Chips* define them as follows: (1) enough observations must be contained in each cell to assure statistical stability; and (2) all of the important explanatory variables must be considered explicitly, in order to avoid false conclusions. Principle (2) was discussed extensively under the cross-classification heading in "Statistical Analysis of Relationships between Variables," elsewhere in this book. There we demonstrated that any variable that is related to both the dependent and independent variables in a cross-classification must be explicitly included in the analysis if false conclusions about

causal relations are to be avoided. The extensive numerical illustrations in the *Wood Chips* reading will serve to reinforce this fact.

Cross-classification need not be limited to simple tabular data, as in the first reading. The article by Coleman, Menzel, and Katz, "Social Processes in Physicians' Adoption of a New Drug," illustrates what is basically a cross-classification approach within the context of a scientific investigation of human behavior. Much of their data was processed extensively before being cross-classified, for example, the "index of simultaneity" plotted in Figures 16 and 17.

The technique utilized throughout the paper is that of comparative graphical analysis, a form of cross-classification. The situation is complicated because the time path of the percentage of doctors adopting the new drug is an important attribute of the data. Therefore, the cells in the classification (that is, Figures 1 and 3-10) contain charts of percentage adoption by months, rather than a single summary number (such as average number of months required to reach 75 percent saturation).

The reader may also note that the analysis used in the Coleman, Menzel, and Katz paper is not a *cross*-classification at all. Each of the charts presents data for two or more values of *one* explanatory variable, for example, number of journals received (Figure 4), or whether or not the doctor attended meetings in his specialty (Figure 5). A true cross-classification would have presented data for all of the possible combinations of variable values, for example, doctors who read six or more journals and attended specialty meetings, those who read six or more journals and did not attend meetings, and so on. Such an approach was clearly precluded here, because of the large number of explanatory variables and the limited number of observations that were available.

Two general questions might receive particular attention as these readings are considered:

1. Is it likely that the effects attributed to explanatory variables are really due to other variables that were not included in the analysis? How did the authors decide which variables should be included?
2. What formal statistical methods, if any, might be used to sharpen up the conclusions drawn from the cross-classification? Under what conditions would their application be worthwhile?

The reader is warned in advance that there probably are no specific "correct" answers to these questions. And yet they are the ones that must be dealt with in the formative stages of any research attempt.

Research Analysis*

.....

A. J. WOOD RESEARCH CORPORATION

SUPPOSE ONE HAS JUST CONDUCTED A MARKETING RESEARCH SURVEY AND has made a preliminary tabulation of the data—the “straight runs.” Some of the data is unexpected, or at least unanticipated, and leads to a number of questions or hypotheses which lead, in turn, to additional analysis, further “cross-runs.” These cross-runs may seem to “explain” the data or resolve the questions that were raised initially; or they may raise further questions, more complicated (or more comprehensive) hypotheses; and again, the cycle is started. New cross-runs are made until we find that we have exhausted our data in the sense that further refinement becomes impossible because our bases “disappear,” that is, the number of people in our sub-groups becomes too small for any meaningful analysis. We may, of course, stop far short of this stage because we have run out of money to pay for the additional analysis; or because we fail to see that additional analysis ought to be made, we fail to see that we have not explained the data we have.

It should be realized, however, that there is considerable danger in terminating the analysis too soon.

Suppose, for example, that we are conducting a survey of shopping behavior in a community which has three supermarkets. We find that 50% of our households shop at X, 30% shop at Y, and 20% shop at Z. All respondents are asked which of the three supermarkets they consider to be “most interested in its customers”: 50% select X, 40% select Y, and 10% select Z.

Now, we might be interested in seeing if those who believe that X is “more interested in its customers” also have a greater tendency to shop at X than those who do not have this attitude. In order to do this, we may set up the following table:

* Reprinted from *Wood Chips*, William Balshem (ed.), Vol. IV, No. 9 (February, 1961).

TABLE 1

<i>Attitude A:</i> <i>X More Interested in Customers</i>	<i>Shop at X</i>		<i>Do Not</i>		<i>Total</i>	
	<i>No.</i>	<i>%</i>	<i>No.</i>	<i>%</i>	<i>No.</i>	<i>%</i>
Hold attitude.....	(275)	55	(225)	45	(500)	100
Do not.....	(225)	45	(275)	55	(500)	100
Total.....	(500)		(500)		(1000)	

We see that among those who hold Attitude A, 55% shop at X; whereas, among those who do not hold the attitude, only 45% shop at X. We may, therefore, be tempted to conclude that holding Attitude A causes respondents to shop at X.

However, a further analysis may turn up the following data:

TABLE 2

<i>Attitude A:</i> <i>X More Interested in Customers</i>	<i>Shop at Y</i>		<i>Do Not</i>		<i>Total</i>	
	<i>No.</i>	<i>%</i>	<i>No.</i>	<i>%</i>	<i>No.</i>	<i>%</i>
Hold attitude.....	(200)	40	(300)	60	(500)	100
Do not.....	(100)	20	(400)	80	(500)	100
Total.....	(300)		(700)		(1000)	

Here we see that among those who hold Attitude A, 40% shop at Y; whereas among those who do not hold the attitude, only 20% shop at Y. By the same reasoning as we used before, we must now argue that holding Attitude A causes respondents to shop at Y.

Now, while in some cases both conclusions might be justified, it hardly seems likely that respondents would shop at supermarket Y because they believe that supermarket X is more interested in its customers. Such might actually be the case: for example, if we asked which supermarket is more "progressive," we might find that some people went to X because they believe it to be a more progressive store (and they want to shop at a more progressive store); while other people shop at Y because they believe X is more progressive (and these people prefer to shop at a more "conservative" store). However, in the case of Attitude A (more interested in customers), such an analysis would not seem to be promising.

Even in the case of Attitude A, however, it may still be a correct analysis. Those of you who are bibliophiles may, for example, recall that many bookstores take special care not to disturb their patrons, but rather allow them to browse undisturbed, bibliophiles being notoriously hostile to any attention paid them by sales personnel. If "more interested in its customers" were interpreted as "more attention paid by sales personnel," we might find many people who would shop at the store where such "attentions" would be frowned upon by the store management.

In any event, it is clear that Tables 1 and 2 present us with a problem and not (necessarily) with a solution. We intuitively feel that something else is at work; that except for "special cases" like our bookstore, people would not be inclined to shop where the management paid them less attention (was less thoughtful) than at some other store; although, again, this might turn out to be the case if those who shopped at Y were agreed that X really did pay more attention to its customers, but that one could get better bargains at Y. In this latter case, of course, our interpretation of Table 2 would have been incorrect: people do not shop at Y *because* they believe X is more interested in its customers but *in spite* of this consideration.

To have terminated the analysis either with Table 1 or Table 2 might have caused us to miss much that is of interest in trying to understand why people do or do not shop at X. This is similar to the problem which arises in the interpretation of statistical findings regarding smoking and cancer. In this latter area, we find that deaths from lung cancer are higher for cigarette smokers than for non-smokers, leading to the conclusion that smoking causes lung cancer. This is similar to our Table 1. However, we also find, and herein lies the problem, that deaths from cancer (other than lung cancer) also are higher for smokers than for non-smokers. Deaths from bone cancer are higher for smokers than for non-smokers. Does smoking cause bone cancer? And deaths from meningitis are higher for smokers than non-smokers: does smoking cause this? Do people shop at Y because X is more interested in its customers? Or is something else at work; have we terminated the analysis too soon?

Another difficulty may arise from quite the opposite direction.

We may, for example, find that there is no higher a proportion of shoppers at X among those who hold Attitude B than among those who do not have this attitude (as in Table 3); nor is the proportion of shop-

TABLE 3

Attitude B:	Shop at X		Do Not		Total	
	No.	%	No.	%	No.	%
Have attitude.....	(200)	50	(200)	50	(400)	100
Do not.....	(300)	50	(300)	50	(600)	100

pers at X higher among those who hold Attitude C than among those who do not have the attitude (as in Table 4).

TABLE 4

Attitude C:	Shop at X		Do Not		Total	
	No.	%	No.	%	No.	%
Have attitude.....	(300)	50	(300)	50	(600)	100
Do not.....	(200)	50	(200)	50	(400)	100

At this point, one may again be tempted not to go further since neither B nor C seems to be related to shopping at X. However, this need not be the case as can be seen in Table 5.

TABLE 5

	<i>Shop at X</i>		<i>Do Not</i>		<i>Total</i>	
	<i>No.</i>	<i>%</i>	<i>No.</i>	<i>%</i>	<i>No.</i>	<i>%</i>
Have Attitudes B and C.....	(120)	60	(80)	40	(200)	100
Do not have Attitude B, do have C. (180)		45	(220)	55	(400)	100
Have Attitude B, not C.....	(80)	40	(120)	60	(200)	100
Do not have Attitude B or C.....	(120)	60	(80)	40	(200)	100

As can be seen in the two upper rows of the table, B shows a definite relationship to shopping at X if we hold Attitude C constant: among those who have C and B, 60% shop at X; among those who have C but not B, only 45% shop at X.

Similarly with the two bottom rows: *among those who do not have C*, only 40% shop at X if they have Attitude B, but 60% shop there if they do not have B.

In the same way, we can see that C is related to shopping at X if we hold B constant. In rows 1 and 3 of the table, we see that *among those who have Attitude B*, 60% shop at X if they also have Attitude C, but the figure is only 40% if they do not. And in rows 2 and 4, we see that among those who *do not have Attitude B*, 60% shop at X if they also do not have Attitude C, but only 45% if they do have C.

Again, to have stopped the analysis with either Table 3 or 4 would have been to miss a very important part of the story.

Let us consider Table 3, again, where we found no effect between Attitude B and shopping at X. Even without considering such a complication as that presented in Table 5, can we say that A does not "cause" X? In the last issue of *Wood Chips*, where we discussed the difficulties in the analysis of an experimental situation, we pointed out specifically the difficulties in the analysis of the "stimulus". . . what the stimulus *is* in contrast to what it is *thought to be*. The following comment illustrates both the difficulty which may be encountered in this stimulus problem as well as the danger of terminating the analysis too soon when negative results (no statistical difference, no significant difference) turn up. The quotation is rather lengthy but, I think, it is instructive and worth quoting in detail.

It is now a long time since I announced an experiment which greatly surprised physiologists: the experiment consists of making an animal artificially diabetic by means of a puncture in the floor of the fourth ventricle. I was led to try this puncture as a result of theoretical considerations which I need not recall; all that we here need to know is that I succeeded at the first attempt, i.e. that I saw the first rabbit on which I operated become strikingly diabetic. But

I afterward had the experience of repeating the experiment many times (eight or ten times) without getting the same result. I then found myself in presence of a positive fact and of eight or ten negative facts; yet I never thought of denying my first positive experiment in favor of the negative experiments which followed it. Thoroughly convinced that my failures were due only to not knowing the true conditions of my first experiment, I persisted in experimenting, to try to discover them. As a result, I succeeded in defining the exact place for the puncture, and showing the conditions in which the animal to be operated on should be placed; so that we can today reproduce artificial diabetes whenever we place ourselves in the conditions known to be necessary to its appearance.

Let me add to the above a reflection showing how many sources of error may surround physiologists in the investigation of vital phenomena. Let me assume that, instead of succeeding at once in making a rabbit diabetic, all the negative facts had first appeared; it is clear that, after failing two or three times, I should have concluded, not only that the theory guiding me was false, but that puncture of the fourth ventricle did not produce diabetes. Yet I should have been wrong. How often men must have been and still must be wrong in this way! It even seems impossible absolutely to avoid this kind of mistake.¹

Now, let us turn to one final example to further illustrate our theme. Suppose we *do* find a difference between some attitude, E, and shopping at Y, as in Table 6.

TABLE 6

Attitude E:	Shop at Y		Do Not		Total	
	No.	%	No.	%	No.	%
Have attitude.....	(300)	60	(200)	40	(500)	100
Do not.....	(200)	40	(300)	60	(500)	100
Total.....	(500)		(500)		(1000)	

Here, again, it may be incorrect to say that E “causes” shopping at Y since the relationship between E and Y may be entirely dependent on some third variable, M, as in Table 7.

TABLE 7

Attitudes E and M:	Shop at Y		Do Not		Total	
	No.	%	No.	%	No.	%
Have E and M.....	(262)	70	(112)	30	(374)	100
Do not have E, have M.....	(88)	70	(38)	30	(126)	100
Subtotal.....	(350)		(150)		(500)	
Have E, do not have M.....	(38)	30	(88)	70	(126)	100
Do not have E, do not have M....	(112)	30	(262)	70	(374)	100
Subtotal.....	(150)		(350)		(500)	
Total.....	(500)		(500)		(1000)	

¹ Claude Bernard, *Introduction to the Study of Experimental Medicine* (Dover, N.Y., 1957), pp. 173-74.

Here we see that when we hold M constant, the relationship between E and Y disappears. We see in rows 1 and 2 that there is no difference in the proportion who shop at Y regardless of whether the respondents have Attitude E or not (among those who have Attitude M). And, similarly, in rows 3 and 4, there is no difference in the proportion who shop at Y regardless of whether the respondents have E or not (among those who do *not* have Attitude M).

We see, therefore, that where we *do not* find any significant relationship between two variables (as in Table 3), an important relationship may lie entangled and undiscovered amid other variables; and where we *do* find a relationship (as in Tables 1 and 6), we may be dealing with one which is spurious or one which only leads to other problems.

Social Processes in Physicians' Adoption of a New Drug*

.....

JAMES COLEMAN, HERBERT MENZEL,
and ELIHU KATZ†

OVER THE PAST 20 YEARS THE PRACTICE OF MEDICINE HAS UNDERGONE profound changes, not the least of which has been the accelerated rate of change itself. Again and again, new diagnostic techniques, new laboratory tests, new drugs, new forms of anesthesia, new surgical procedures, and new principles of patient management make their appearance. Most of these innovations are minor steps which alter the medical scene only by gradual accretion, if at all; many, indeed, are short-lived or quickly superseded; a few have been milestones in medical developments. Whatever the ultimate significance of a new practice may be, its immediate fate, once it has been launched from the laboratory or research clinic, rests in the hands of the practicing physician in the field.

The pathways by which a successful innovation in medical practice spreads through the profession have seldom been systematically investigated. The present study is a contribution toward that end. It concerns the fate of a single innovation in four cities, a new variant in a well-established family of drugs. Such a case study can hardly claim to represent the course of all types of medical innovations under all circumstances, but it does illuminate some important paths and processes through which medical innovations can make their way into the practice of physicians. One may add that drugs are peculiarly suitable as tracers of these

* Reprinted from the *Journal of Chronic Diseases*, Vol. IX, No. 1 (January, 1959), pp. 1-19. Received for publication from the Bureau of Applied Social Research, Columbia University, on September 26, 1958.

This is the first of two articles and may be identified as Publication No. A-273 of the Bureau of Applied Social Research, Columbia University. Preparation was facilitated by funds provided by a grant to the Bureau of Applied Social Research from the Eda K. Loeb Fund.

Philip Ennis, Marjorie Fiske, Rolf Meyersohn, and Joseph A. Precker participated in the planning of the study. Helmut Guttenberg and Sydney S. Spivack gave invaluable aid in the statistical analysis and in the collection of data, respectively.

† Johns Hopkins University, Columbia University, and the University of Chicago.

paths, because they are physical objects with standardized names, their release dates are easily ascertainable, and pharmacists maintain exact prescription records.

The data of this study stem from two sources. In four midwestern communities, interviews were conducted with 125 general practitioners, internists, and pediatricians, who generously contributed an average of $1\frac{1}{2}$ hours of their time. They constituted 85 per cent of all practitioners of their specialties in these cities. In addition, data on the prescriptions written by these 125 physicians were obtained from almost all the pharmacies in these communities.¹ The prescription data covered a period of 16 months beginning with the release date of a new drug which will here be called "gammanym." At this time two older drugs of the same general type were in widespread use. They had appeared some years earlier and are here designated as "alphany" and "betany." The three medications belong to a well-established family of drugs which has widespread applicability in the hands of general practitioners and many specialists. By the end of the survey period, 16 months after its release, gammanym had become at least a part of the standard medication of most practicing physicians; 87 per cent of the general practitioners, internists, and pediatricians in the sample under study had introduced it into their practice. But this change had been neither immediate nor all-encompassing. The over-all rate of introduction of gammanym by these doctors is seen in Fig. 1. This cumulative curve indicates for each stated date the percentage of doctors who had already prescribed the drug up to that time. As can be seen, more and more doctors introduced the drug during this 16-month period until almost 90 per cent had used it; then the curve finally leveled off.

To be sure, the rapid rise in the use of gammanym did not mean that its predecessors, alphany and betany, were dropped from use. The degree of overlapping use of two or three of the drugs by the same doctor is pictorially represented in Fig. 2, for three time intervals, representing the beginning, middle, and end of the period studied. Thus, for example, as late as 16 and 17 months after the release of gammanym,² only 22 per cent of the doctors were prescribing gammanym exclusively, while 15 per cent were prescribing both betany and gammanym, 13

¹ Ninety-one additional interviews were held with practitioners of other specialties, but their prescription record was not examined. The communities had altogether 356 physicians in active private practice. The sample was designed to include all the general practitioners, internists, and pediatricians, and a selected group of other specialists. The population of the four cities ranged from about 25,000 to just over 100,000.

² The prescription data were 3-day samples at intervals averaging $28\frac{1}{2}$ days, which are here termed "months," so that the 16-month survey period contains 17 such "months." Because of the nature of time sampling, the introduction dates recorded here may be somewhat later than the physicians' actual introduction of the drug.



FIGURE 1. Cumulative proportion of doctors introducing gammanym over a 16-month period ($N = 125$).

per cent were prescribing both alphanylm and gammanym, and 20 per cent were writing prescriptions for all three drugs during these same 2 months (more accurately, during the 6 sampling days representing these 2 months). The introduction of gammanym meant for most doctors an addition to whatever drugs of this type they were already using, rather

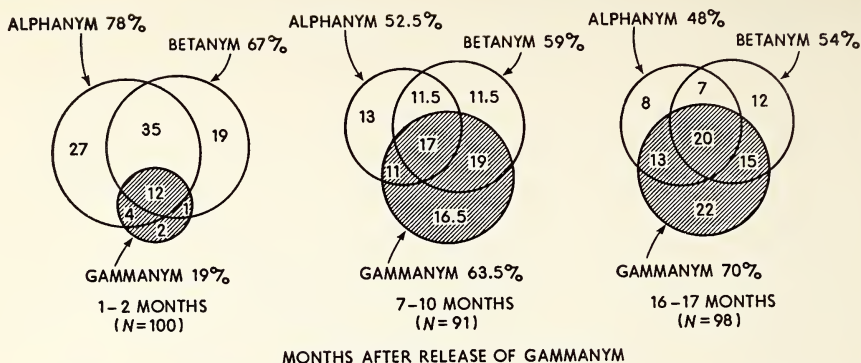


FIGURE 2. Overlapping use of alphanym, betanym, and gammamym during three time intervals.

than a substitution. Only slowly, if at all, did some of the doctors stop using the older drugs.

Drug Adoption and Individual Characteristics of Physicians

This, then, describes the over-all course of the acceptance of gammamym in these four communities, from the time it first arrived on the drugstore shelves to the time when it appeared on prescription slips of most of the doctors in town. But what exactly happened during this period? Through what paths and processes did the new drug find its way into the prescribing habits of the local physician? As one essential step in answering these questions, it is necessary to examine and compare the doctors who introduced the new drug quickly with those who prescribed it only after most of their colleagues had done so. By ascertaining the backgrounds, types of practice, and other characteristics of these doctors, we can begin to sketch a picture of the "innovator" among the local practicing physicians.

Let us first look at specialty differences. Pediatricians introduced gammamym into their practices more quickly than internists, and internists introduced it more quickly than general practitioners, as Fig. 3 shows. The average pediatrician first prescribed the drug 6.6 months after its release, the average internist 8.2 months, and the average general practitioner 9.0 months after release. This result is somewhat surprising in view of the general impression that internists are the pace setters. But the contradiction is only apparent and results from the different prescription volume of the different specialties. The pediatricians' average number of prescriptions for alphanym, betanym, and gammamym combined was 13.6 per 3-day sampling period; for internists it was 2.7, and for general practitioners 3.6. When doctors with roughly the same volume of prescriptions are compared, the contradiction disappears. Inter-

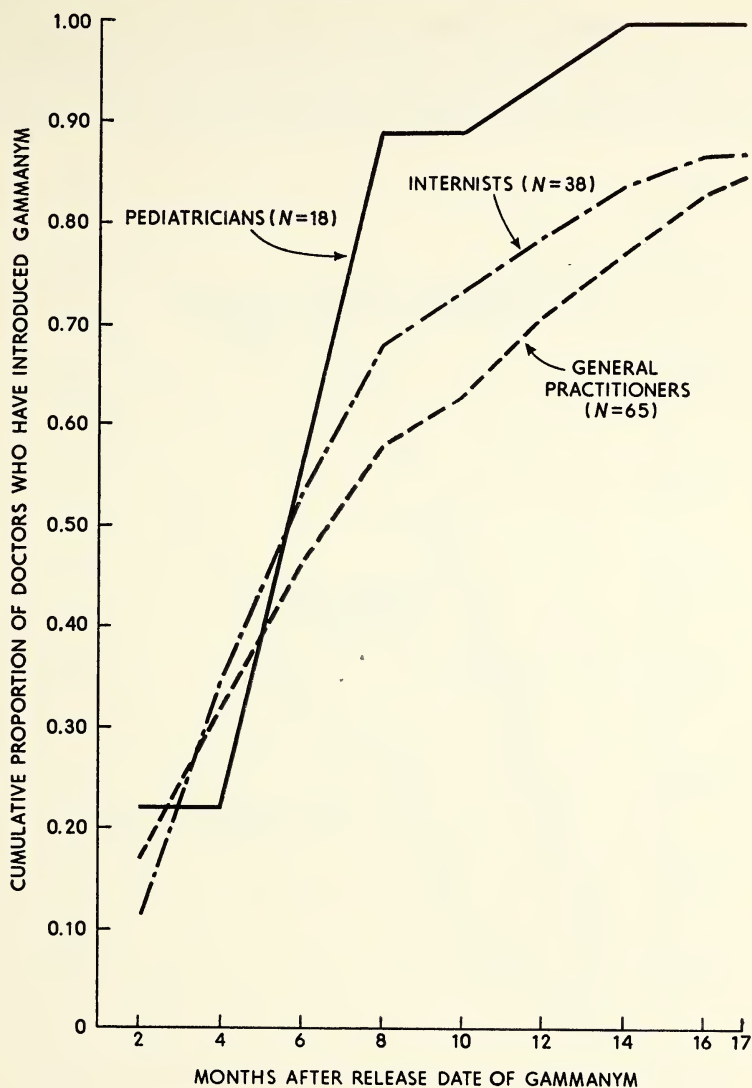


FIGURE 3. First use of gammanym by specialty.

nists then introduce the new drug more rapidly than pediatricians or general practitioners (not shown).³

Less surprising is the finding that doctors who expose themselves frequently to the primary sources of information were much more likely to be innovators than those who do not. Figs. 4, 5, and 6 show, respec-

³ A more detailed and comprehensive account of our research results is being prepared for publication by The Free Press under the tentative title *Doctors and New Drugs*. Selected aspects are treated in detail in certain articles. (See the first four references at end of article.)

tively, the more rapid rate of gammanym introduction among those who receive many journals, among those who attend many out-of-town specialty meetings, and among those who conscientiously attend conferences in their own hospitals. It is difficult to assess to what extent this is due

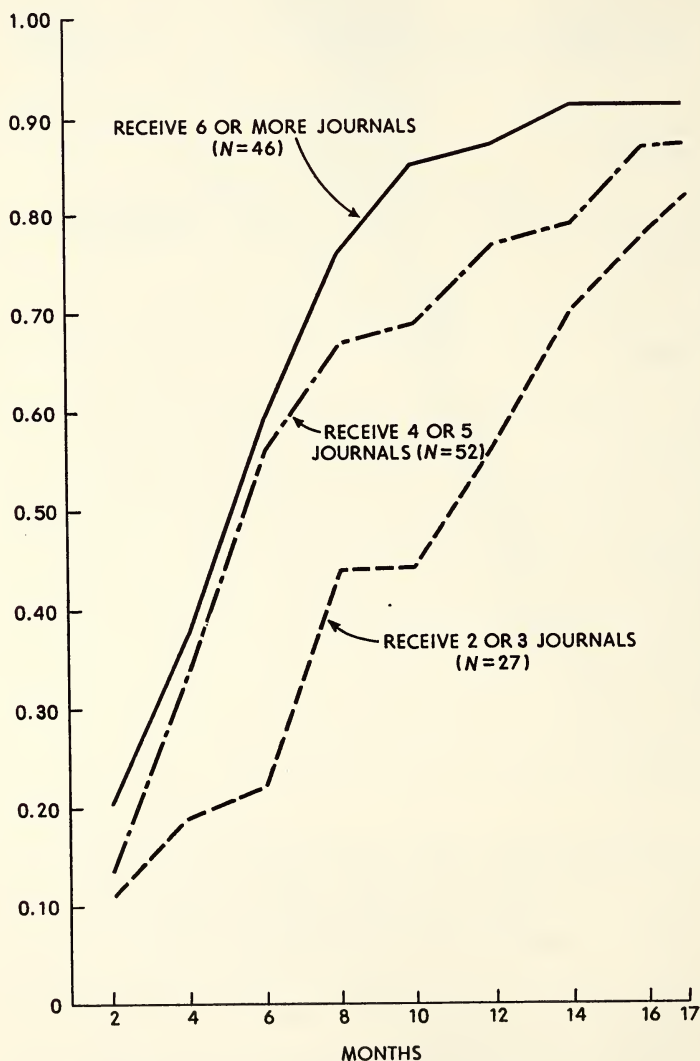


FIGURE 4. First use of gammanym by number of journals received.

to an actual effect of these means of communication on the doctor's prescription habits. Quite possibly it means merely that doctors who are sensitive to new developments read more and go to more meetings. What we do know is that the relationship of drug introduction to journal reading and meeting attendance is independent of the doctor's specialty.

When doctors in each specialty are considered separately (not shown), the findings of Figs. 4, 5, and 6 remain essentially unchanged. In fact, reading and attendance, as shown in Figs. 4, 5, and 6, have a considerably stronger relationship to early use of gammanym than does the doctor's

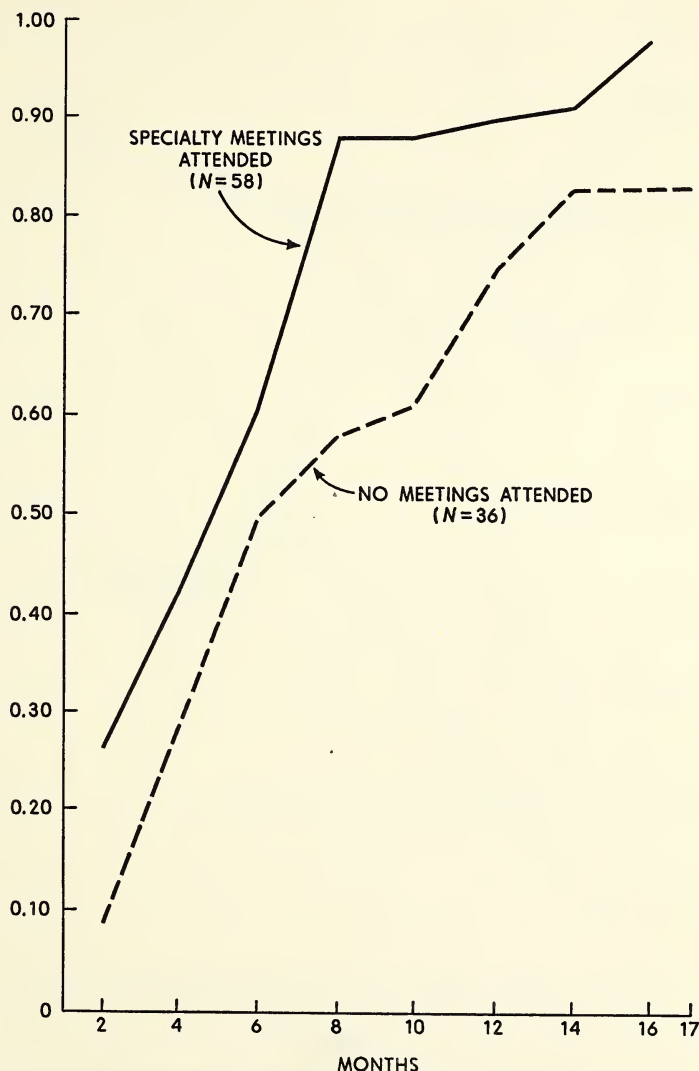


FIGURE 5. First use of gammanym by specialty meetings attended.

specialty (Fig. 3). It appears that the innovator is less characterized by his specialty than by voluntary activities like attendance at meetings and reading journals which bring him into closer contact with events in the profession.

But this is not equally true for *all* information-getting activities. There

are some potential sources of influence or information, exposure to which shows no relationship to early use of gammanym. Figs. 7, 8, and 9 show that doctors who read many pharmaceutical house organs, those who attend many nonspecialty meetings (American Medical Association,

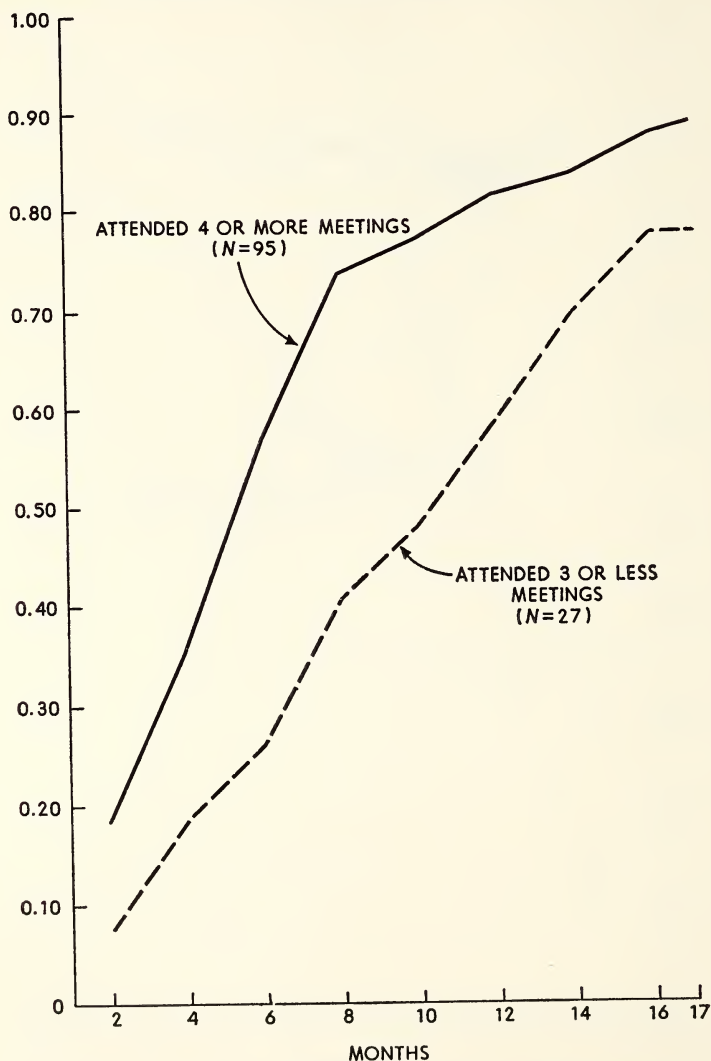


FIGURE 6. First use of gammanym by hospital meetings attended.

state and regional medical societies, etc.) and those who attend county medical society meetings regularly were little or no quicker to introduce the drug than their colleagues who attend those media less regularly. These media apparently attract a fairly representative audience of doctors among whom sensitivity to new developments is not more prevalent than among physicians in general.

Drug Adoption and Physicians' Contacts with Colleagues

The media discussed so far have been the relatively obvious channels through which innovations may be diffused to local doctors. But doctors

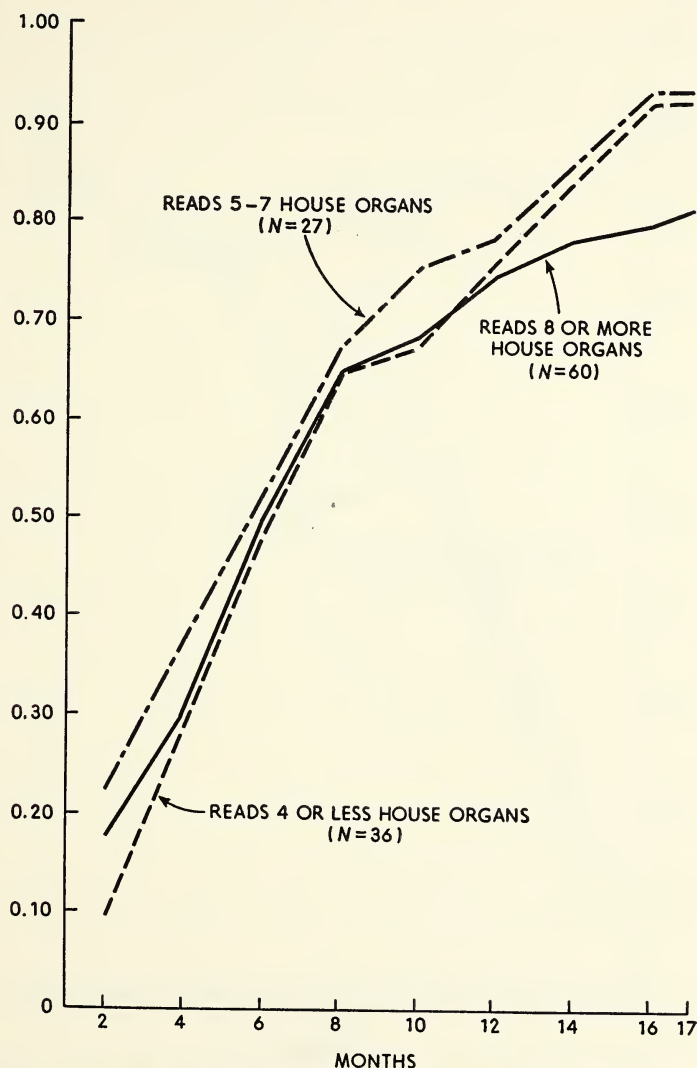


FIGURE 7. First use of gammanym by number of house organs read.

were also found to be affected in their drug adoptions in some less obvious ways. One path which might not have been anticipated is suggested by Fig. 10, the office arrangement of the doctor. By simply dividing doctors into those who share offices with one or several colleagues and those who have an office alone, we find a considerable difference. The

doctors who share offices introduced the drug an average of 2.3 months sooner than their colleagues who practice alone.

It is useful to ask just what social and psychologic processes may produce the effect shown in Fig. 10. Two hypotheses seem quite reasonable.

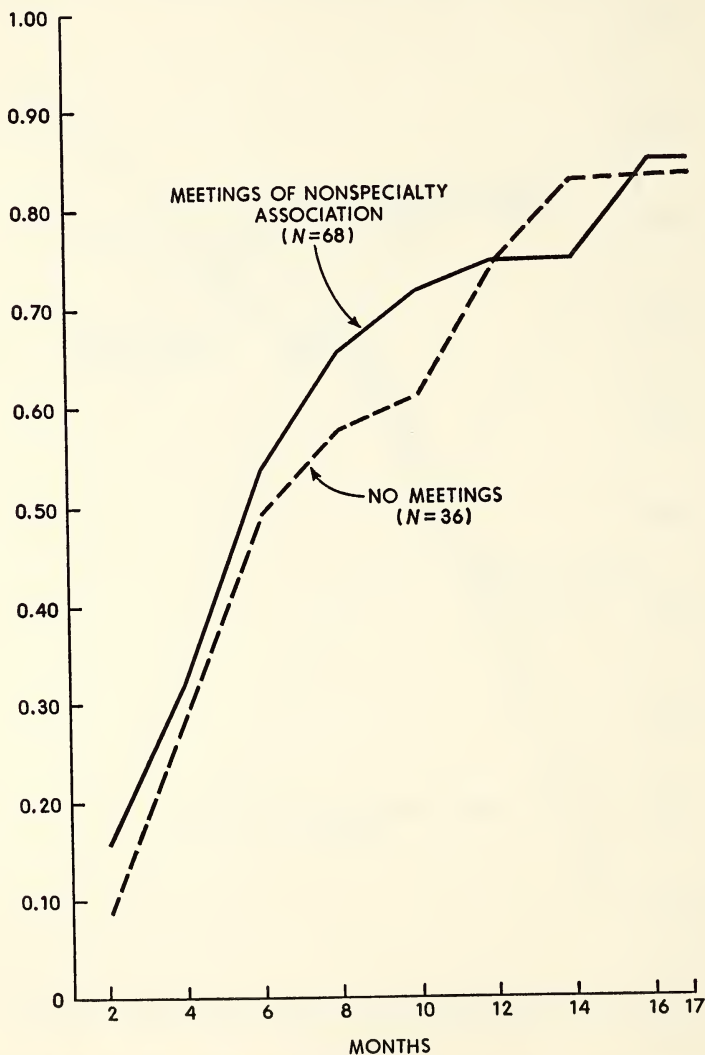


FIGURE 8. First use of gammanym by out-of-town nonspecialty meetings attended.

First, being in close professional contact with colleagues keeps a doctor well informed, so that he is saved the difficulty of finding out for himself about each new development. He has surrogates to carry part of the burden of finding out about new developments, for, as soon as any of his office partners seriously consider trying a new therapy or even just find out about a new development, they will discuss it with him.

A second interpretation of the results, however, seems equally reasonable. Introducing a new technique into his practice is always somewhat dangerous for the physician. He has no first-hand knowledge of possible ill effects, yet he must shoulder the blame if ill effects should

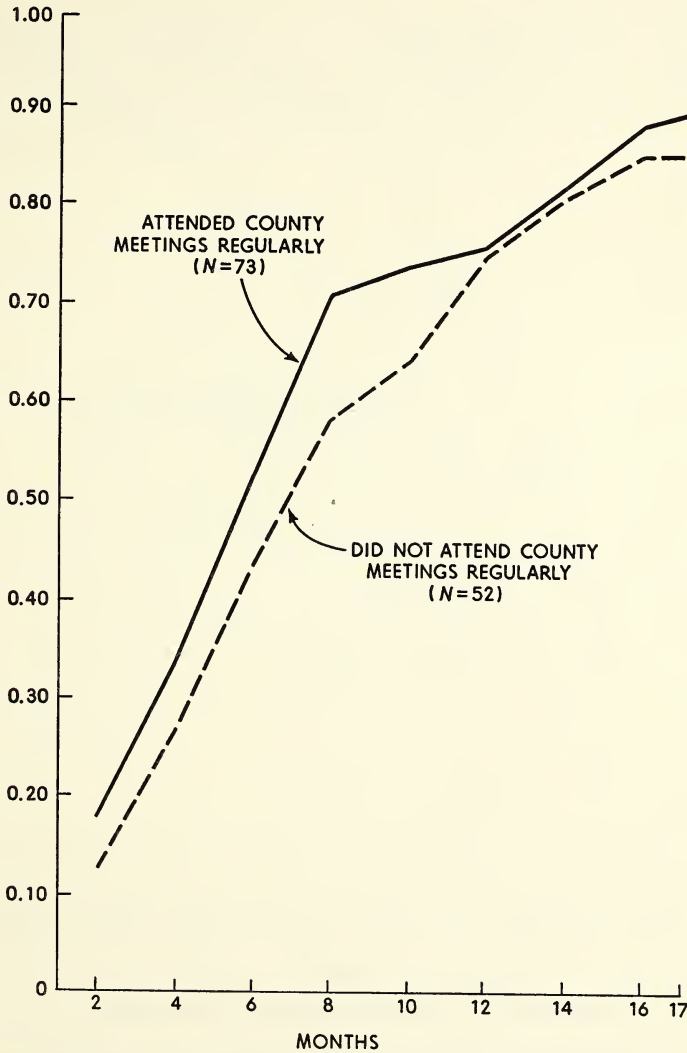


FIGURE 9. First use of gammanym by attendance at county medical society meetings.

occur. Because of this, the doctor needs all the reassurance he can get from his fellows to lessen the uncertainty which he faces. The doctor who shares an office with others can, in a sense, depend upon their support and use it for reassurance, while the doctor who practices alone must make the bold step without this added support. It is difficult to determine which of these two interpretations is most valid without

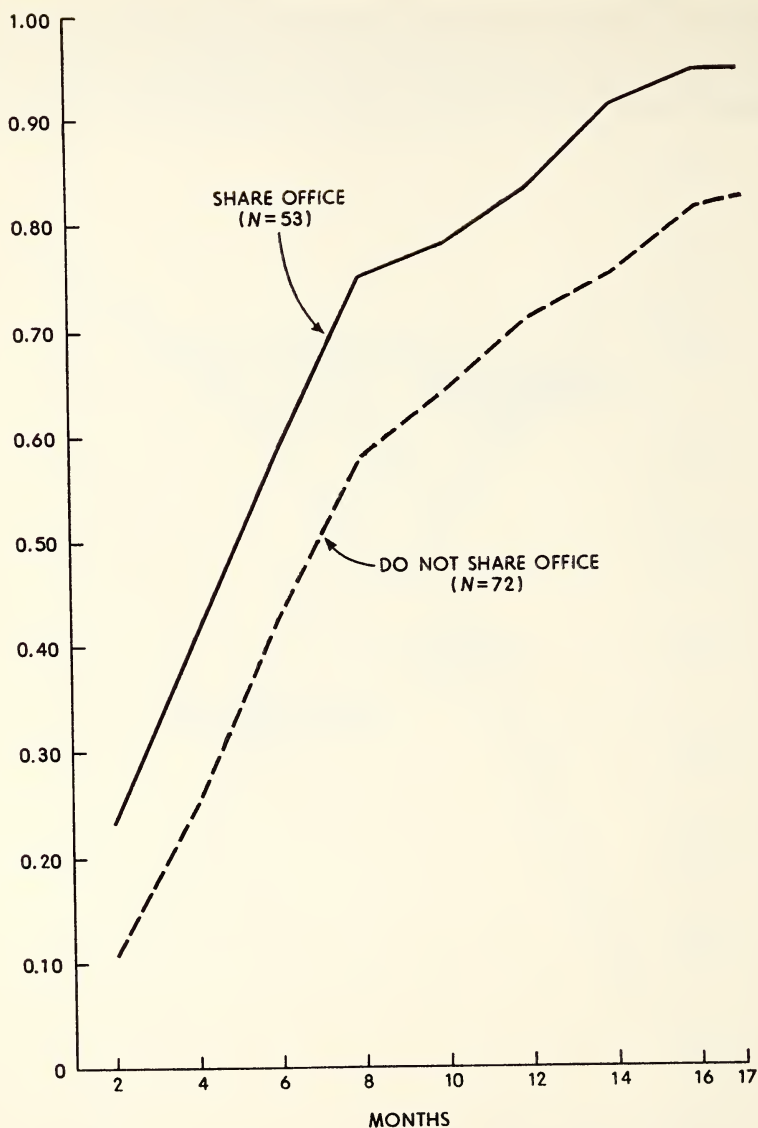


FIGURE 10. First use of gammanym by office sharing.

knowing whether the primary barrier to drug innovation among doctors who are alone is lack of information or lack of assurance.

But for whatever reason office-sharing leads to earlier use of new drugs, it is apparent that a doctor's social location in the community of his local colleagues affects his drug use by giving him access to information, by providing him with assurance, or in some other way. This leads to a more general question. What about the dozens of other relations that a doctor may have with his colleagues, beside that of office partnership?

What, for example, might be the importance of his contact with colleagues in the hospital during leisure hours? Would it affect his use of new practices and products? In order to answer questions like these, one must know the relations among the doctors of the community, that is, one must know the social structure of the medical community which these relations make up. To uncover the most important facets of this structure, three questions were asked of each doctor in the interview: (1) "When you need information or *advice* about questions of therapy, where do you usually turn?" (2) "Who are the three or four

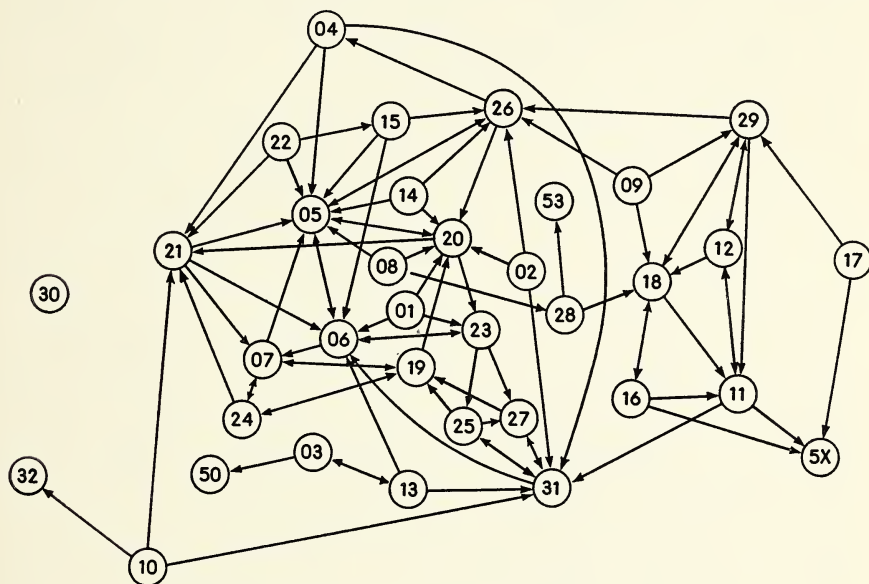


FIGURE 11. Discussion network in City D.

physicians with whom you most often find yourself *discussing cases* or therapy in the course of an ordinary week—last week, for instance?" (3) "Would you tell me who are your three *friends* whom you see most often socially?" In answer to each of these questions, three names of other doctors were requested. The replies to the three questions yielded three cross-sections of the structure of the medical communities, much like a biologist's sections of a plant fiber along several axes. Fig. 11 presents, as an illustration, one of these cross-sections for one of the four cities covered in the survey. This particular diagram pictures that city's structure of discussion partnerships, that is, a sectioning according to the replies to the second question above. In this sociogram, as such a diagram is called, each circle represents a physician, who is identified by a code number. An arrow pointing from Circle 04 to Circle 05 means that Dr. 04 named Dr. 05 as one of his most frequent partners in the discussion of cases. The double-headed arrow connecting Circle 05 and Circle

06 means that Dr. 05 and Dr. 06 each named the other as a frequent partner in the discussion of cases. The fact that 7 different arrows point to Circle 31 (near bottom) means that 7 different colleagues named Dr. 31 as a frequent discussion partner; in other words, Dr. 31 is quite popular as a discussion partner.

Our general question now becomes how does this social structure of the community of physicians in a city facilitate or inhibit the diffusion of a new drug? This is an immensely complex question to answer—as complex as the social structures themselves. But it is possible to break the structure and the question into simpler components and to investigate them one by one. Perhaps the simplest question that can be asked in this connection is what is the difference in the rate of drug adoptions between doctors who have contact with many and with few colleagues? This question is a direct parallel to the earlier question about the difference in the rate of drug adoption between doctors who have office partners and those who have none. One can ask more specifically what is the difference in the rate of drug adoptions between doctors who are named as advisors (or as discussion partners, or as friends) by many of their colleagues and those who are not named by any? One may call the first group socially integrated, and the second group socially isolated.

We have pointed out that doctors with office partners introduced gammanym more quickly than those who practice alone. This leads to the hypothesis that the well-integrated doctors would be quicker to introduce this new drug than their more isolated colleagues. This is indeed what is found to be true, and it is true for each of the three cross-sections of the social structure which are under examination here, as shown in Figs. 12, 13, and 14. The effect is quite strong in the predicted direction. Doctors highly integrated into each of the structures were much quicker to introduce the new drug than the more isolated doctors. The average doctor among those most frequently named as advisors introduced the new drug 3.1 months before the average of those who are never named as advisors. The corresponding mean difference between those integrated and isolated as discussion partners is 4.1 months, and between those integrated and isolated as friends is 4.3 months. The importance of these factors can be gauged from the fact that no other factor examined in the entire study yielded a mean difference of more than 4 months, with the single exception of total volume of prescriptions for this general type of medicine. These results suggest that the networks of informal relations among doctors were highly effective as chains of information and influence in the diffusion of this innovation.

Yet, one might reasonably object, is it not likely that these differences in integration and isolation merely reflect different personality characteristics among these doctors and that it is really these personality differences, and not the contacts with other doctors, which account for the striking results found? This is especially plausible in the case of the ad-

visorship network. After all, the doctors who are often designated as advisors may very well be chosen by their colleagues precisely because they are aware of new medications. In that case their early introductions of new drugs would be a cause rather than a consequence of their high

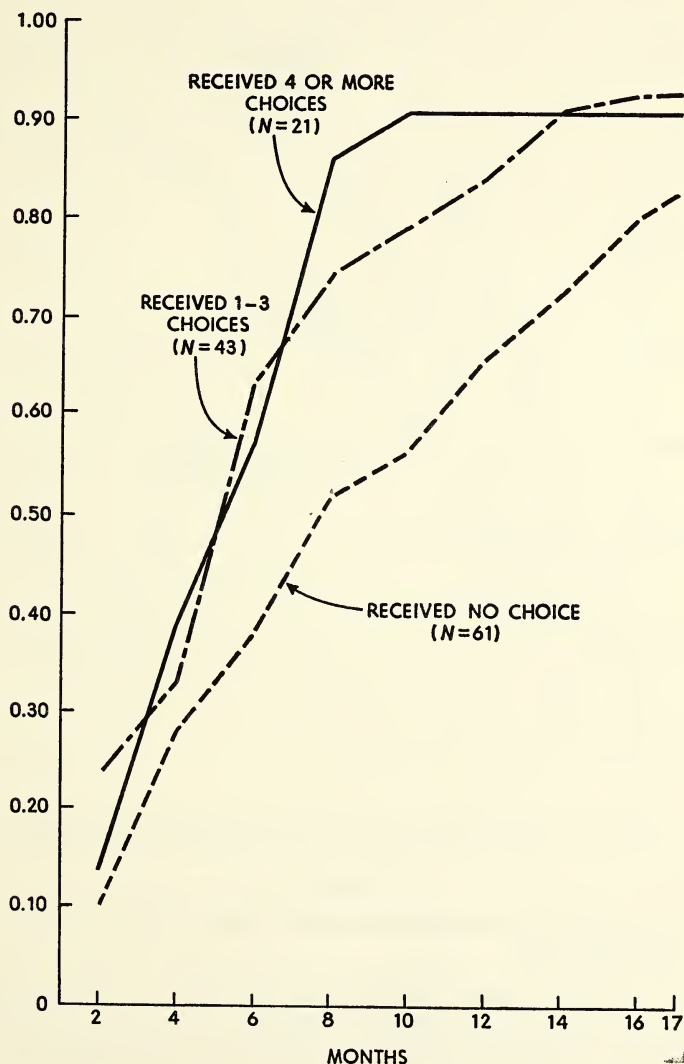


FIGURE 12. First use of gammanym by number of choices received as advisor.

integration. Yet, contrary to this interpretation, the effect of the advisorship network is smaller, not larger, than that of the networks of discussion partners and friends. It is the friendship network which yields the largest difference (4.3 months) between the average adoption times of the integrated and isolated doctors. Since an "integrated" doctor here

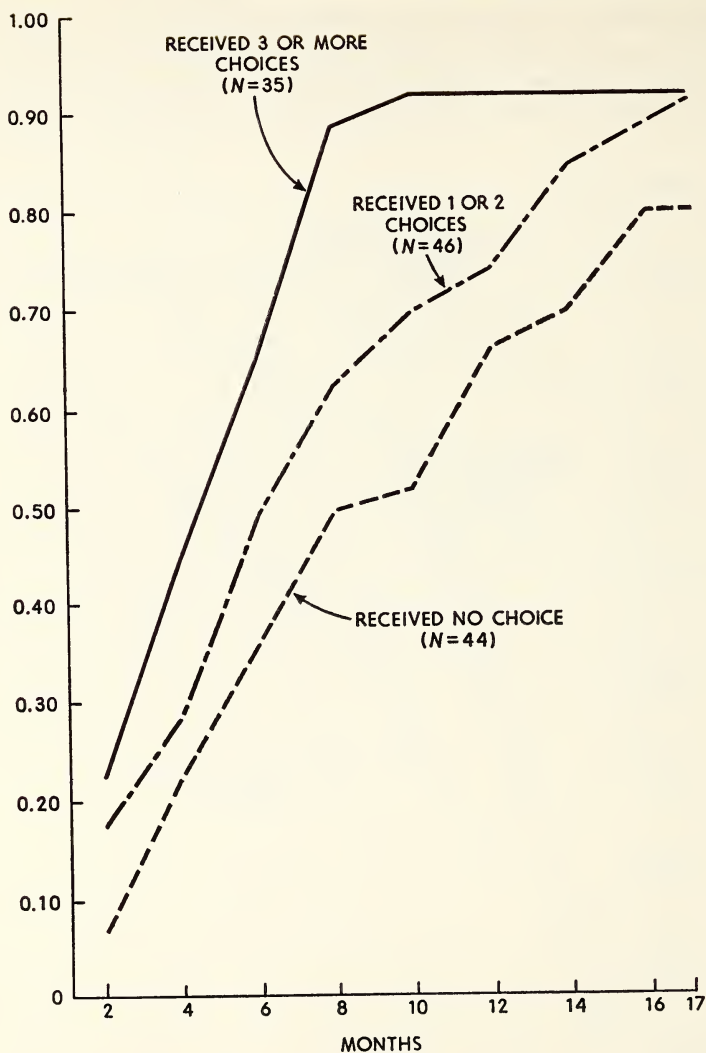


FIGURE 13. First use of gammanym by number of choices received as discussion partner.

means one named by many colleagues as a "friend . . . seen most often socially," it is very unlikely that his integration is a result of some professional habit which is accompanied by early introduction of new drugs. It is much more likely that early introduction of the new drug was conditioned by the doctors' informal contacts with one another, through the network of friendships, and through the more strictly professional relationships as well.

Two Processes of Diffusion

But even stronger evidence for this claim can be found in another

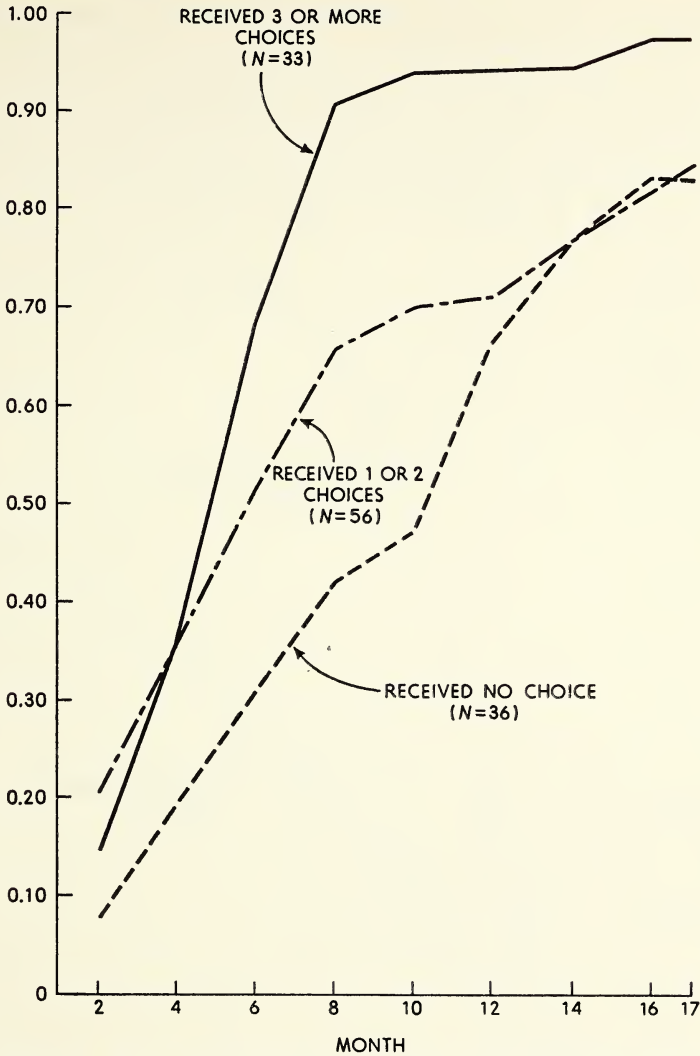


FIGURE 14. First use of gammanym by number of choices received as friend.

place. If the socially integrated doctors tended to introduce the drug early merely because of some personality characteristic, the use of the new drug should have spread among these integrated doctors in very much the same way as among the isolates, except earlier. If, on the other hand, it is true that the socially integrated doctors owe their early introduction of gammanym to the networks of contacts which surround them, then the use of the new drug should not only spread *earlier* among them than among the isolates, but the very *nature of the process* of diffusion should then be different. At the extremes, there would be these two processes of diffusion: (a) Among the isolated doctors, it would be an individual process. The effective stimuli—such as detail men, medical journals, ad-

vertising from drug houses, and so on—remain fairly constant throughout the diffusion period. The number of doctors introducing the new drug each month would remain a constant percentage of those who have not already adopted the drug. A typical graph for this process would look like the lower curve of Fig. 15. (b) Among the integrated doctors, it would be an interpersonal or “snowball” process. If, for example, one pioneer introduces the new drug and converts a colleague to

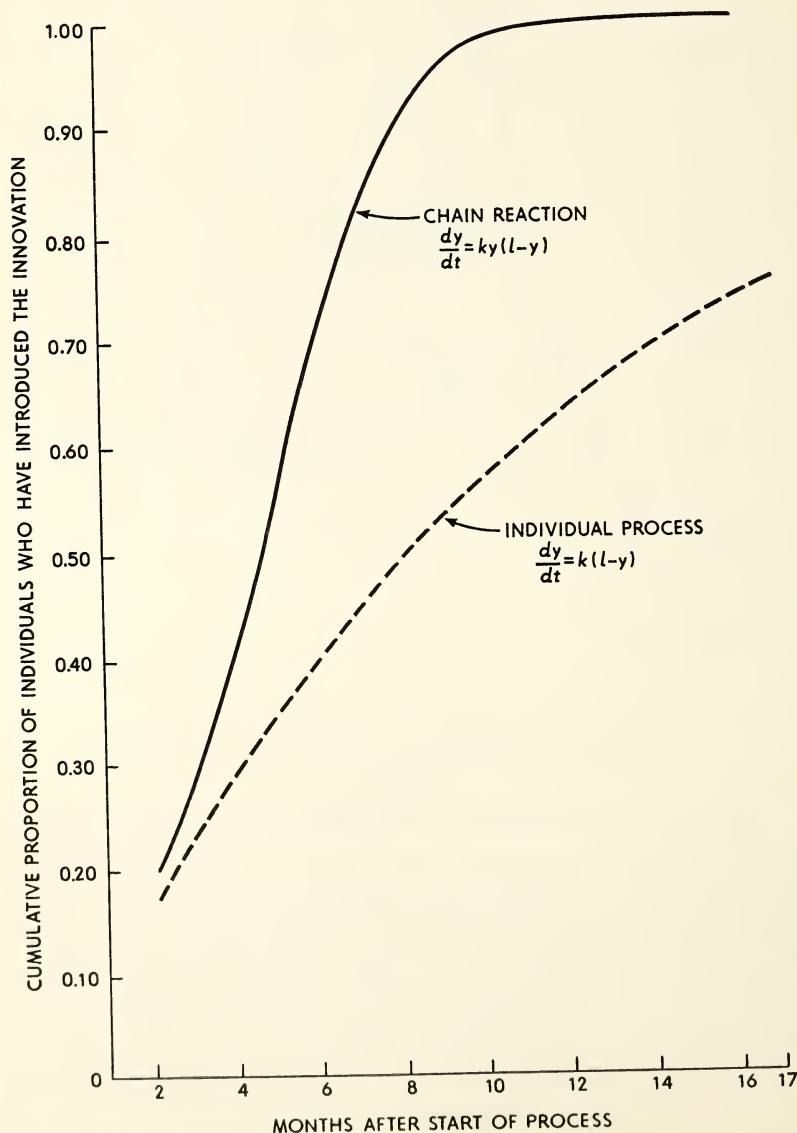


FIGURE 15. Comparison of a model of chain-reaction innovation with a model of individual innovation.

it during the first month, then these two doctors will convert two others during the second month; during the third month, there would be four doctors making new converts; and so on. The number of doctors introducing the new drug each month would not remain a constant percentage of those yet to be converted, but would gain headway in proportion to those who have already been converted. A typical graph for this process would look like the upper curve in Fig. 15.⁴

Note that the two *theoretical* curves in Fig. 15 differ not only in average height, but also in shape. The curves start out with the same proportion of initial users and then diverge sharply. The curve for the interpersonal process turns upward or "snowballs" in response to the proportion who already use the drug. The curve for the individual process bends gradually downward, indicative of response to constant stimuli. This difference in shape corresponds closely to the differences between the *empirical* curves for the extreme groups in each of Figs. 12, 13, and 14. In each instance, the curves start out with roughly the same proportion of users in the first 2 months and then diverge sharply. The curves for the integrated doctors continue steeply upward, indicating at least in part a person-to-person process. The curves for the isolated doctors bend gradually downward, as would be expected in the case of an individual diffusion process. To be sure, none of these empirical curves fit exactly the extremes shown in Fig. 15.⁵

Figs. 12, 13, and 14, which showed the relationship of adoption rates to the doctor's contacts with his colleagues, can now be contrasted with Figs. 3-6, which showed the relationship of adoption rates to a series of individual characteristics (specialty, number of journals read, and so on). The upper and lower curves in each of Figs. 3-6 often diverged sharply from the beginning, and the upper curves had essentially the same shape as the lower curves. This is as expected, if the doctors represented by these upper and lower curves differ in individual receptivity to or awareness of innovations, but not in their location in effective networks of interpersonal relations.⁶ While these results constitute only a case study of the diffusion of one new product among the doctors of four cities,

⁴ This process, well known in the study of rates of reaction in chemistry and in population studies, is known as the logistic law. It has been spoken of as the "Law of Social Change," e.g., Ridenour (in reference list).

⁵ A detailed examination of the correspondences and contrasts will be undertaken (reference 1). More concise numerical documentation of the finding is reported elsewhere (reference 3).

⁶ It is worth noting that office partnership (Fig. 10), a "social" attribute, shows generally the same shape as the "individual" variables of Figs. 3-6. Office partnership, however, constitutes a different kind of social configuration than do the networks represented in Figs. 12-14. A doctor who shares an office with another is in contact with *that* other doctor and not part of an interlocking network. Thus while he may benefit from his partner's sensitivity to the new drug and thus come to adopt it earlier, he is not a part of a chain reaction or "snowballing" system by virtue of this partnership.

they have important implications. They suggest that a doctor's tendency to innovate is not only a function of something about him as an individual, but also—and more strongly—a function of his social location among other doctors. The social and professional contacts a doctor has with his colleagues evidently serve important functions in the diffusion of a new practice, which are not duplicated by journals, meetings, detail men, or drug house advertisements.

But this result leaves much unsaid. What is it about the social networks that affects the innovating behavior of men within them? Is it merely transmission of information, or do these networks provide the doctor in an unclear situation with the security of numbers? And what are the various stages of the diffusion process as the new drug changes from one used by a small minority into one used by almost the whole community? These are the questions to be considered now.

The Question of Simultaneous Adoptions by Doctors Who Associate with Each Other

How can some of these more complex questions about the stages of diffusion be examined? One way, though certainly not the only way, is to examine pairs of doctors who maintain some specified form of contact with one another. Each doctor, it will be recalled, had been asked three questions about his relations to his colleagues. (To whom did he turn for *advice*? With whom did he most often *discuss cases*? What *friends* did he see most often socially?) A doctor and any colleague whom he named in reply to any of these questions constitute a pair of related doctors.

If the networks of doctor-to-doctor contacts are effective, then pairs of related doctors should be more alike in their behavior than pairs assorted at random. More specifically, if a chain-reaction process of drug introduction is at work, then, it seems, adjacent links in the chain—that is, pairs of related doctors—should introduce the drug at about the same time, ideally, during the very same months. If, on the other hand, the use of gammanym was not transmitted through these networks, then the interval between the gammanym adoption dates of doctors and their advisors, doctors and their discussion partners, and doctors and their friends would be no shorter than those of any two doctors picked at random.

Actually, the average intervals between the dates of gammanym adoption of each doctor and those whom he had named as his advisors, discussion partners, or friends was almost identical to the average intervals for pairs picked at random.⁷ This meant the rejection of our original hypothesis that pairs of doctors in contact would introduce the drug more

⁷ The "random" formula used makes allowance for the earlier gammanym adoption of the more integrated doctors. For details on the statistical procedures, see Coleman, Katz, and Menzel (reference 3).

nearly simultaneously than pairs of doctors assorted at random. There was, on the other hand, the earlier evidence that the doctor's integration was important to his introduction of gammanym. This dictated a more intensive look at the behavior of pairs of doctors. Accordingly, we raised the question whether the networks, though ineffective for the *whole* period studied, may have been effective for the *early* period, immediately after the drug was marketed. This, indeed, proved to be the case.

The Stages of Social Diffusion

In order to describe this tendency more precisely, an *index of simultaneity* has been devised, constructed separately for each month. As applied here, it measures how closely the drug introduction of doctors during a given month followed the introductions by any associates who

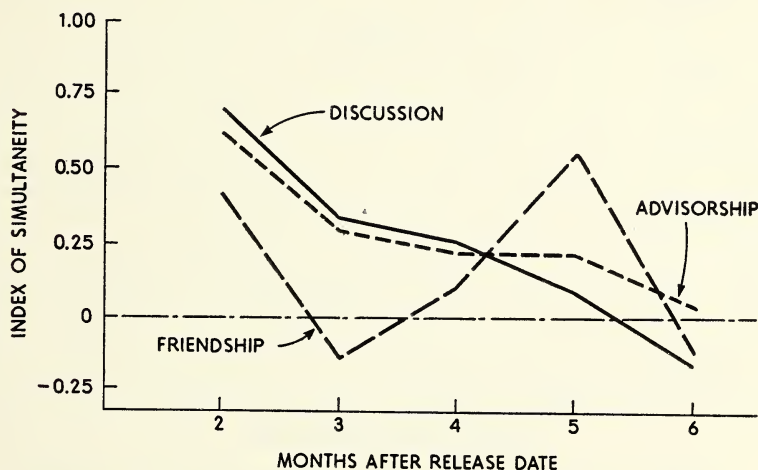


FIGURE 16. Index of pair-simultaneity for three networks at different times.

had adopted the drug during the same month or earlier. This index would have a value of +1 if each doctor had introduced the drug during the same month as his discussion partner (or friend, or advisor); it would be zero if the introduction dates of the two doctors in a pair were as far apart as expected by chance.⁸ In Fig. 16, the values of this index are plotted. Separate curves are plotted for pairs of friends, pairs of discussion partners, and advisor-advisee pairs. Comparing the three structures, it appears that the discussion network and the advisor network are much alike in their effects. Both are most effective during the earliest period. Both are somewhat more effective than the friendship network

⁸ The index automatically compares the actual closeness to a random model and also makes allowance for the fact that, for example, adopters during the first month could not possibly follow anyone by more than one month, while later adopters could. For further discussion of the index and the statistical methods used, see Coleman, Katz, and Menzel (references 1 and 3).

during these early periods. The friendship network, on the other hand, appears to have its maximum effectiveness later, about 5 months after gammanym appeared on the market. Finally, after about 6 months, none of the networks any longer show an effect.

These findings, which show different effectiveness of the social networks at different times after the drug's release, have several implications which will be examined shortly. But first, the very structure of the networks has its own implication. Whatever effect these networks have should operate first in the more dense parts of the structure, where a number of lines of social relationship converge and should only then spread out to the more open parts of the structure, that is, to the relatively isolated doctors. It has already been shown that the more isolated

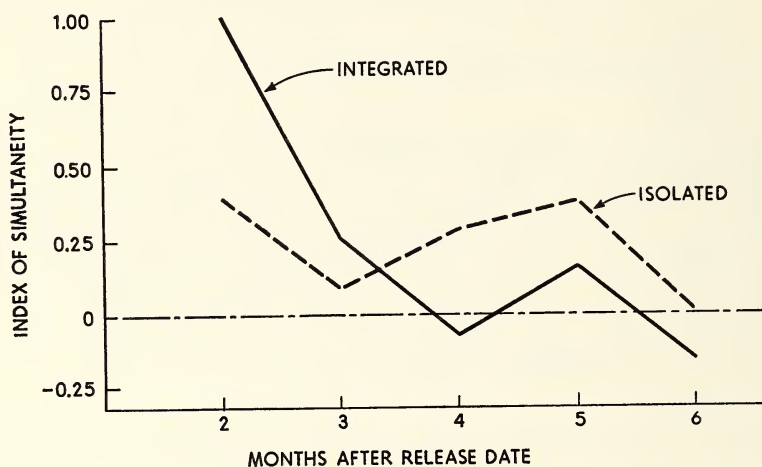


FIGURE 17. Index of pair-simultaneity at different times for doctors differing in integration.

doctors, on the average, introduced gammanym considerably later than the socially more integrated doctors. We propose, however, that when more isolated doctors did introduce the drug early, it was not with the help of the social networks. While the networks were operative as channels of influence early for the integrated doctors, they were operative only later for the more isolated ones. This is what seems to have occurred. Fig. 17 plots the index of simultaneity separately for more and less integrated doctors. (The graphs show weighted averages for all three networks; separately the numbers of cases would be so small as to produce erratic trends.)

The peak of effectiveness of doctor-to-doctor contacts for the well-integrated doctors appeared in the earliest month for which it can be plotted (the second month) after which effectiveness sharply declined. For the relatively isolated doctors, by contrast, the networks were not so effective at first as were those for the integrated doctors, but they main-

tained their effectiveness longer. Thus it appears that the networks of relations were effective not only for the more integrated doctors but also for those relatively isolated doctors who introduced the drug during the first 5 months of the drug's availability.

To summarize the results so far, it appears that social influence in the process of drug adoption occurred in several stages. First, it operated only among the doctors who are most integrated into the community of their colleagues through ties of a professional nature (as advisors or as discussion partners). Then it spread through the friendship network to doctors who are closely tied to the medical community through their friendship relations. By this time, social influence had also become operative in the more diffuse parts of the social structure, that is, among the relatively isolated doctors. Finally, there came a phase during which an occasional doctor still introduced gammanym, but in complete independence of the time at which his associates introduced it; the networks now showed *no* effect. For the integrated doctors, this phase began 4 or 5 months after the drug's release. For the isolated doctors, it began about 6 months after the drug's release. By this time, the social structure seems to have exhausted its effect. Doctors who introduced gammanym into their practices after this time apparently responded exclusively to influences outside the social networks, such as the professional journals, detail men, drug house advertisements, and so on. They did not, it appears, depend upon their personal relations with other doctors for information and influence. The channels of influence between doctors had operated most powerfully during the first few months after the release of the new drug. Such influence as any doctor had upon his immediate associates by his introduction of the drug occurred very soon after the drug became available. Why is this?

The Role of Contacts with Colleagues in Clear-Cut and Ambiguous Situations

One answer is that it is only in the early months after a drug's appearance that a doctor needs the support and judgment of his colleagues. It is chiefly when the drug is new that the doctor who is to adopt it needs his colleagues to confirm his judgment and to share the feeling of responsibility in case the decision to adopt the drug should be wrong. At this time, familiarity with the new drug is minimal and the doctor is in an uncertain situation. Several sociopsychologic experiments have shown that it is precisely in situations which are objectively unclear, situations in which the individual's own senses and other objective resources cannot tell him what is right and what is wrong, that he needs and uses social validation of his judgments most fully.⁹ The first months following the release of a new drug appear to present just this kind of situation. To be sure, the particular drug innovation with which this re-

⁹ See reference 6.

search has dealt was not a dramatic one. Presumably the effects shown here would be much stronger in the case of a more radical innovation. It is nevertheless suggested that the reason for the greater effectiveness of contacts with gammanym users during the earliest months after the release of the drug was due to the greater uncertainty about the drug that prevailed at that time. It is proposed that a doctor will be influenced more by what his colleagues say and do in uncertain situations than in clear-cut situations.

This hypothesis also implies that the necessity for support and validation of judgments by colleagues would be much greater in decisions about some medical conditions than about others. It would be great where the physiology of the illness is not well understood and the treatment is subject to much trial and error. It would be small in conditions which are well understood and in which the action of the medication is well known. It is possible, in fact, to test this implication with certain data of this study. This time the data do not refer to the time of introduction of a new drug, but rather to the habitual use or nonuse of certain classes of modern drugs for two specified conditions. These were respiratory infections and mild-to-moderate cases of essential hypertension. Respiratory infections allow few alternatives of treatment, and their success or failure becomes quickly apparent. They present a clear-cut situation. Hypertension, on the other hand, allows many kinds of treatment and their success can be gauged only slowly and with difficulty. Hypertension presents an uncertain situation.

The implication of the general hypothesis is therefore that pairs of related doctors should be more alike (compared to chance expectations) in their treatment of hypertension than in their treatment of respiratory infections. In order to test this hypothesis, the doctors were classified first according to whether or not they named a broad-spectrum antibiotic as "the antibiotic or sulfonamide [they] most commonly used in infectious conditions of the respiratory tract." They were also classified according to whether or not they included *Rauwolfia serpentina* drugs in their "usual treatment for essential hypertension in mild or moderate cases." Finally, the *Rauwolfia* users were divided according to whether they preferred reserpine or other *Rauwolfia* preparations. With respect to the first classification, a doctor and his friend were considered alike if both or neither named a broad-spectrum drug; with respect to the second, doctors were considered alike if both or neither included *Rauwolfia* drugs; and, with respect to the third, they were considered alike if both or neither preferred reserpine to other *Rauwolfia* drugs.

Table I gives the results in terms of an index of homogeneity. This index would be +1 if all pairs were alike; it would be zero if pairs of related doctors were no more often alike than expected by chance; and it is negative if pairs of related doctors are less often alike than expected by chance. The results confirm the hypothesis: in all three networks, pairs

TABLE 1

HOMOGENEITY OF PAIRS OF RELATED DOCTORS IN TREATMENT OF RESPIRATORY INFECTIONS AND HYPERTENSION*

	<i>Advisor- Advisee Pairs</i>	<i>Discussion Pairs</i>	<i>Friendship Pairs</i>
Respiratory infections (broad-spectrum versus other antibiotics).....	-0.056	-0.016	0.029
Hypertension (Rauwolfia included or not)...	0.054	0.095	0.109
Hypertension (reserpine versus nonres- erpine).....	0.280	0.410	0.233
N (number of pairs).....	(253)	(258)	(211)
	(165)	(151)	(121)
	(111)	(97)	(71)

* These data refer to all 216 interviewed physicians, not only to the 125 whose prescription record was examined. Cf. footnote 1.

are homogeneous beyond chance expectations in hypertension treatment but hardly at all in the treatment of respiratory infections. This finding, like that of the simultaneity of gammanym adoptions by pairs of doctors in the early months, presumably arises from the need for social support and social validation in situations where authoritative objective validation is scant.

SUMMARY

Two hundred and sixteen physicians in four cities were interviewed, and the prescription records of 125 general practitioners, internists, and pediatricians among them were searched, in order to study how the use of a new drug termed "gammanym" spread through these communities of physicians. The main findings are:

1. Gammanym was introduced earlier by doctors with a large volume of prescriptions for this general type of drug, by those who exposed themselves frequently to certain media of information, and by those who shared their offices with one or several partners.

2. Doctors who maintained a variety of contacts with a large number of colleagues, the socially integrated doctors, typically introduced the new drug into their practices months before their relatively isolated colleagues. The degree of a doctor's integration was measured by the number of his colleagues who named him, in response to certain interview questions, as an advisor, frequent discussion partner, or frequently visited friend.

3. Among the integrated doctors, the use of the new drug spread at an accelerating rate, indicating an interpersonal process of diffusion, while among the isolated doctors use of the new drug spread at a constant rate, indicating largely individual responses to constant stimuli outside the community of doctors.

4. The hypothesis that a doctor and his friend, a doctor and his advisor, and a doctor and his discussion partner would tend to introduce the new drug at about the same time was not borne out for the period as a whole.

5. During the early months following the drug's release, however, doctors who introduced the drug tended to follow closely upon any associates who had adopted it earlier.

6. This phenomenon was strongest during the very earliest months in the

case of pairs of discussion partners and advisor-advisee pairs. In the case of pairs of friends, it reached its peak about 2 months later. In all three cases, the phenomenon occurred among the relatively isolated doctors as well as among the integrated doctors, but it reached its peak much later in the case of the isolated doctors.

7. The apparent greater effectiveness of contacts with colleagues during the early months was attributed to the greater uncertainty about the new drug that prevailed at that time. This interpretation is supported by comparisons of uncertain and clear-cut situations of another sort; pairs of related doctors were found to be more alike in the drugs they use for essential hypertension than in the drugs they use for respiratory infections.

A second article will inquire into the possible relationship of early gammanym introduction by a doctor to a more general tendency to be receptive to innovations.

REFERENCES

1. COLEMAN, JAMES; KATZ, ELIHU; and MENZEL, HERBERT. *Doctors and New Drugs* (tentative title). Glencoe, Ill.: Free Press of Glencoe, Inc. (in preparation).
2. MENZEL, HERBERT, and KATZ, ELIHU. "Social Relations and Innovation in the Medical Profession: The Epidemiology of a New Drug," *Public Opinion Quarterly*, Vol. XIX (1955), p. 337.
3. COLEMAN, JAMES; KATZ, ELIHU; and MENZEL, HERBERT. "The Diffusion of an Innovation among Physicians," *Sociometry*, Vol. X (1957), p. 253.
4. MENZEL, HERBERT. "Public and Private Conformity under Different Conditions of Acceptance in the Group," *Journal of Abnormal and Social Psychology*, Vol. LV (1957), p. 398.
5. RIDENOUR, LOUIS N.; SHAW, RALPH R.; and HILL, ALBERT G. *Bibliography in an Age of Science*, Urbana, Ill.: University of Illinois Press, 1951, pp. 30-35.
6. SHERIF, MUZAHER. "Group Influences upon the Formation of Norms and Attitudes," in Maccoby, Newcomb, and Hartley (eds.), *Readings in Social Psychology*. New York: Henry Holt, 1958, pp. 219-33.

Regression and Discriminant Analysis

REGRESSION AND DISCRIMINANT analysis are important techniques for quantifying the characteristics of causal relationships, which are inferred "by eye" in cross-classification studies. Their purpose is to measure a numerical relationship between a particular dependent variable and one or more explanatory, or independent, variables. In regression analysis, both the dependent and explanatory variables are measured in numerical terms. Discriminant analysis is designed for use where the explanatory variables are numerical, but the dependent variable is dichotomous, for example, "yes" or "no," "black" or "white," and so on.

Regression analysis has long been an important tool for economists in their attempts to measure aggregate consumer demand. It is only recently that the procedure has come out of the economist's kit of tools and into that of the practicing marketing analyst. Examples of both economic and marketing applications are presented in the readings of this section.

The article by Juréen, "Long-Term Trends in Food Consumption—A Multi-Country Study," is representative of a broad class of economic demand studies; Juréen's own work in this area goes back to the 1930's. His study was aimed at determining the effects of income and prices upon demand for a number of different food categories. Beginning with a series of scatter diagrams, he estimated the income coefficient of the demand function for each product category by means of a univariate regression analysis. Then multiple regression was used to estimate the effect upon demand of the food category's own price, and those of each of the other categories. It is interesting to note that income was not included in the price effects equation: its value was held constant so as not to bias the estimates of the price elasticities. Among the regression

problems explicitly considered by Juréen are (1) the form of the assumed regression relationship; (2) the effects of the existence of errors in variables as well as in equations; and (3) the effect of multicollinearity (defined by him as a functional relationship between the explanatory variables) upon estimates of structural parameters.

Seymour Banks' article, "Some Correlates of Coffee and Cleanser Brand Shares," illustrates the possibilities and difficulties of applying the regression technique to marketing problems. The reader will want to distinguish between the kinds of variables that are important for Juréen and Banks. Those used by Juréen are broad aggregates of variables whose behavior is fairly well defined—although estimates of their magnitudes may be in error. Banks, on the other hand, dealt with disaggregated data about particular brands, practices in particular groups of stores, and actions of particular consumers. While some of his variables are well defined, others had to be represented as *ad hoc* "indices" (for example, indices of stock display, promotional effort, and point of purchase advertising).

The analysis of detailed data, such as that used by Banks, is an essential part of the marketing picture; it deals with the kinds of relationships that are of interest to the business firm. We must note also that the marketing analyst must be prepared to quantify variables that are not normally measured in quantitative fashion. It is generally better to get a crude estimate of a variable's magnitude into the analysis rather than to neglect its effects entirely; to do this is to assume that it has no effect or that its effect is small, and this may be the only assumption that is known to be false.

The reader should also focus upon the way in which the fundamental assumptions of regression apply to Banks' analysis. His sample sizes are extremely small: 9 brands for cleanser and 21 for coffee. Therefore, his tests of the significance of regression parameters depend upon the normality of the error term in the regression equation (this assumption is not required for large samples). Can Banks' sample be regarded as drawn by random selection? Are any variables excluded from the regression equation which are likely to be highly correlated with both market share and any of the independent variables? How highly are the independent variables correlated among themselves? (Banks does not present data on multicollinearity. How would this affect our assessment of his conclusions?) Does the fact that the sample size equaled only 9 have any bearing upon our interpretation of the fact that the multiple correlation coefficient was very high? Finally, we should examine Banks' view about the relationship between regression analysis and experimentation.

Franklin Evans' article, "Psychological and Objective Factors in the Prediction of Brand Choice—Ford versus Chevrolet," gives us an example of the use of multiple discriminant analysis in a marketing context. His goal was the prediction of Ford and Chevrolet ownership in a sample

of respondents, by using a number of psychological and socioeconomic measures as explanatory variables. The article has a significant bearing upon the efficacy of motivational versus traditional marketing research techniques, but our concern here is with the statistical aspects of the work.

Evans' problem is representative of a broad class of dichotomous brand choice situations. It is a "natural" for discriminant analysis. Information on 10 personality and 14 socioeconomic variables was included in the analysis. The means of each variable for Ford and for Chevrolet owners, together with the multiple discriminant function weight, are presented separately for each set. The socioeconomic variables gave a slightly better prediction of brand ownership than did personality measures, but neither function was able to discriminate very effectively. The linear discriminant function containing a combination of 13 psychological and socioeconomic variables did little better than those for the sets separately.

Evans concludes that his results do not reveal an effective discrimination between Ford and Chevrolet owners. This would tend to imply that the populations owning Ford and Chevrolet automobiles have similar distributions of the psychological and socioeconomic characteristics studied. If there really is very little difference between the two populations, as would appear to be evidenced by the scatter diagrams presented in the Evans' paper, no statistical technique can do anything but reproduce this fact. The linear discriminate function, properly applied and interpreted, is the most efficient procedure for attempting to discriminate between dichotomous variables, given a particular set of data.



Some Correlates of Coffee and Cleanser Brand Shares*

SEYMOUR BANKS†

A THEORY OF MARKET DEMAND FOR BRANDS MUST CONSIDER TWO MAJOR elements: first, the choice process within the mind of the consumer; and second, the marketing environment in which purchase takes place. This paper describes a model of market demand for brands of convenience goods and reports the results of a test of this model.

All discussion of demand in this paper is in terms of ratios, i.e., a brand's share of the market. If one attempted to deal with demand in an absolute sense, these ratios would have to be multiplied by a base which would consider such factors as the importance of the product in consumers' budgets and the level of national income. This is a task far greater than seems desirable at the moment, and one which is not necessarily required for realism. Many businesses consider primary demand trends to be out of their control, and evaluate their relative success in terms of selective demand position.

The general demand model may be written:

$$P_i = f_c(A_1, A_2, \dots) + f_r(B_1, B_2, \dots) + f_w(D_1, D_2, \dots) + f_m(E_1, E_2, \dots)$$

C_i
 R_i
 W_i
 M_i

where P_i is a brand's share of the market. Market share is taken to mean a brand's share of the total volume of sales of the given product class in a certain geographical area.

The terms on the right of the equation are of two types. The first term (C_i) deals with consumer evaluation of the intrinsic attributes of a brand, and the remainder with the marketing efforts of the component

* Reprinted from the *Journal of Advertising Research*, Vol. I, No. 4 (June, 1961), pp. 22-28. Copyright by the Advertising Research Foundation, Inc., 3 East 54th Street, New York 22, New York.

† Leo Burnett Company, Inc.

elements of the channel of distribution: R for retailer, W for wholesaler and M for manufacturer.

The A 's in the consumer term of the above equation are criteria by which consumers evaluate the intrinsic qualities of various brands of a given type of merchandise. For coffee, these qualities might include flavor, flavor consistency, bouquet, type of grind, and size and type of package. These criteria will differ from person to person in number and importance. Furthermore, since judgment is subjective, individuals with identical criteria may have different evaluations of a given brand. The evaluation of each brand on all criteria considered by a consumer leads to a consolidated judgment of that brand at that time.

This mental evaluation of brands by a consumer can be visualized as an archipelago with some peaks rising out of a sea while others are visible below the surface. Sea level corresponds to a level of acceptability for brands of a given product. The peaks represent scalar evaluations of the qualities of the various brands at a given time. Brands are considered acceptable in the sense that, by their intrinsic qualities alone, they would be considered as possible purchases. For example some brands of coffee may not be acceptable because they have too mild or too strong a flavor or do not come in the desired grind.

But the above picture holds only temporarily. As time passes, brands may lose their acceptability, either because their qualities have actually deteriorated or because other brands have been improved. Brands previously unacceptable may rise to acceptability by product improvement. A scouring cleanser, for example, which was changed from an abrasive to a detergent increased sales considerably.

Then too, the *level* of acceptability is subject to change. In times of shortage, consumers take almost any brand. But in a buyer's market, they will not accept substitutes for favored brands.

A purchase is made from among the acceptable brands but is not mechanically determined by an evaluation of value, either ordinal or cardinal. An acceptable product may cease to be bought because the customer who used it previously desires a change for change's sake. This satiation phenomenon appears to be random and is relevant only for individual decision; its effect probably washes out in groups (Banks, 1950).

The number of brands considered depends in part upon the extent of the consumer's experience and in part upon the nature of the product. The more experienced the consumer, the more brands he knows, but limits are imposed by attention and memory. Generally the consumer is more familiar with convenience goods than with shopping goods. In the case of shopping goods like appliances, a complete picture of brands is seldom available—the consumer shops not only to learn which brands are for sale, but often to discover the criteria by which he might evaluate the brands he has discovered.

The term of the demand equation starting with R represents selling effort by retailers for each brand considered. The B 's represent their performance of activities like special displays, demonstrations, recommendations to consumers, services rendered (large stock, credit and repair), return privileges, etc., on the brands he carries. The next term deals with efforts of wholesalers to push different brands, training courses for retailer salesmen, demonstrations, special price or credit concessions and so on. Finally, we have the term which represents the selling efforts of the manufacturers for each of their brands.

The rather simple assumptions implied by the form of equation used are not really satisfactory in representing the effect of a manufacturer's sales efforts. A sound marketing program calls for working at several levels simultaneously. Manufacturers merchandise new consumer advertising campaigns to wholesalers and retailers; retailers are affected by advertising campaigns addressed to the general public; price changes affect margins throughout the channel of distribution. The manufacturer's selling efforts and those of his wholesalers and retailers are often closely related. Because of this, the plus signs in the equation should be interpreted as general logical conjunctions rather than as arithmetical additions. Possibly multiplication signs would represent reality more closely.

Customers, retailers, wholesalers and manufacturers vary greatly in scope of activity and our equation must not be interpreted as giving equal weight to each of the terms. Formally, the f_c , f_r , f_w and f_m in the equation represent quite general functions of the factors inside the brackets.

METHOD

Two research techniques which are often used to determine the effect of marketing variables upon sales of brands are experimentation and regression analysis. In the first, the researcher controls the way in which the independent variables affect his test units. One example might be a sales test of a new package design versus an old, each package being used in a comparable sample of stores.

In the regression procedure, the researcher assumes a simple relationship between the market share of a brand and prices, promotional efforts, point of purchase advertising, etc., for it. The assumption is that the market share of a brand can be expressed as an equation, usually linear, which is of the form:

$$\begin{aligned} \text{Brand share} &= a (\text{price of the brand}) \\ &+ b (\text{consumer's preference rating}) + \text{etc.} \end{aligned}$$

By mathematical techniques we choose values of the coefficients a , b , c , etc., which best fit the observed facts. The researcher cannot control factors such as preference ratings for all the brands, but his mathematical procedures enable him to estimate the effect of each while the effects of

the others are accounted for statistically. The regression procedure has administrative advantages, but will not yield the functional relationship among the variables studied.

The user of regression analysis assumes that:

1. The values of the independent variables are fixed and may be looked upon as population parameters. Often particular values are deliberately chosen.
2. For a given set of values of the independent variables, the resulting values of the dependent variable are normally distributed.
3. That the sample be drawn by a process of random selection (Anderson and Bancroft, 1951).

The research situation which we shall discuss has an additional complication in that all variables, both independent and dependent, are subject to error. Bartlett (1949), and more recently Acton (1959), have discussed procedures for dealing with this type of situation, but few applications have appeared in the literature.

In evaluating the results of tests of significance of regression coefficients, caution will be used. In a numerical example presented by Bartlett (1949), the 95 per cent confidence interval of the regression coefficient is 16 per cent larger, assuming both variables are subject to error, than when assuming the independent variable is stated or measured without error.

Our data were collected in early December 1950 from 165 Chicago housewives selected by area sampling procedures. Blocks were chosen at random and four respondents picked at random in each block. Purchases were measured about a week before information was collected on other variables but it was felt that the situation prevailing at the time of purchase could not differ materially from that a week later.

For both scouring cleanser and coffee, the interview covered the respondent's knowledge and use of the various brands, preference ratings on the brands she knew, and brands, quantities, and place of her last purchase. Information was also obtained on possibilities for exposure to advertising in terms of ownership of a radio or TV set, subscription or regular readership of magazines and Chicago newspapers. Respondents were classified into high, medium and low economic strata on the basis of the 1940 rent data for the block in which they lived; this last rating was subject to revision by the interviewer after inspection of the household furnishings and equipment.

The respondents were asked to state their preferences for brands of scouring cleanser and coffee by means of a thermometer rating device. One-half of the respondents made preference statements before the question of purchases was raised and the other half made similar preference statements after the interviewer had determined the brands on hand. The two preference distributions differed insignificantly; from which we in-

ferred that no bias was induced by the order of questioning, and the returns of both halves were combined.

Only the highest preference ratings made by respondents were considered. If a respondent placed several brands in the highest category she used, each brand received an equal fractional share of the rating. The sum of these ratings for all respondents gave the total number of highest-choice ratings per brand. This procedure is described elsewhere (Banks, 1950).

To obtain data on purchases, the interviewer asked to see all containers of the last purchase under the guise of obtaining code numbers. Only when the containers were reported destroyed (e.g., when coffee packed in bags had been put into canisters and the bag discarded) were housewives asked to tell what brand they had bought last. Brand shares were of total amount bought in last purchases.

After discussing brand preference and purchase with a respondent, it was easy to discover where her last purchase of scouring cleanser and coffee had been made. The interviewer then went to the designated store or stores (not infrequently the scouring cleanser and coffee were bought in different stores) as soon as possible after finishing a given block assignment of interviews, and for each of the brands carried observed the price (in cents per package for scouring cleanser, in cents per pound for coffee), the amount of stock displayed, and the presence of promotional effort and point of purchase displays.

The formal model of demand discussed at the beginning of this article must be simplified drastically for empirical research because it deals with a very large number of variables, most of which are extremely difficult to measure. The model became, after appropriate simplification, one of multiple linear regression. The equation on the next page was set up to study the forces affecting market shares of brands of scouring cleanser and coffee.

Information on consumer advertising expenditures was obtained from three sources. A. C. Nielsen Company made available (in private correspondence) radio, newspaper and magazine advertising expenditures for brands of scouring cleanser and coffee in metropolitan Chicago from June through November 1950. This was satisfactory for scouring cleanser but gave no information on advertising of chains' private brands of coffee. The *Chicago Tribune* made available unpublished data on total advertising expenditures of these chains in Chicago in newspapers during this period. Some chains were willing to state, also in private correspondence, what share of their local advertising budget was allocated to their private brands of coffee; for the others, a sample of newspapers was selected and the ratio of space found to be allocated to their brands of coffee was used as the share of their total advertising budget allocable to their private brands of coffee.

$$P_i = aX_1 + bX_2 + cX_3 + dX_4 + eX_5 + fX_6 + gX_7,$$

where

P_i = Each brand's share of the sample's last purchase of the product.

X_1 = Consumer preference in terms of number of highest ratings per brand.

X_2 = Average price in cents per unit.

X_3 = Store coverage =

$$\frac{\text{No. stores stocking each brand, weighted by number shopping these stores.}}{\text{Total number of users}}$$

X_4 = Index of stock display =

$$\frac{1 (\text{no. good shelf displays}) + 2 (\text{no. special displays}) \times 100}{\text{Total number of ratings}}$$

X_5 = Index of promotional effort =

$$\frac{\text{No. stores where brand carried offers of price deals or premium} \times 100}{\text{No. of stores stocking}}$$

X_6 = Index of Point of Purchase Advertising =

$$\frac{\text{No. stores where brand had POP effort} \times 100}{\text{No. of stores stocking}}$$

X_7 = Dollar expenditure for advertising in the three major media (newspaper, radio and magazines), Chicago, June through November 1950.

a, b, c, d, e, f, g are the regression coefficients which were computed mathematically.

RESULTS

The data were analyzed to determine first, how successfully—as measured by the coefficient of multiple correlation—the research model fitted the actual purchase pattern; and second, the relative importance of the different elements of the model, as measured by the size of their regression coefficients.

First we considered how closely each factor separately was related to brand shares. For this we examined the simple correlations. For both coffee and scouring cleanser, consumer preference rating and store coverage, themselves highly correlated, showed the highest simple correlation with market shares. For scouring cleanser, promotional effort was highly correlated with market shares, while advertising expenditure was poorly correlated. The reverse was true for coffee. For both products, advertising expenditure was more highly correlated with store coverage than with market share or any other variable.

Table 1 shows the regression coefficients which permit direct evaluation of the relative effect of the independent variables on the dependent variable, brand shares.

For the scouring cleanser equation, all coefficients were significant at the five per cent level of confidence. The most important factors in de-

TABLE 1

REGRESSION COEFFICIENTS BETWEEN BRAND SHARES AND SEVEN MARKETING ACTIVITIES

Marketing Activity	Cleanser ($N = 9$)	Coffee ($N = 21$)
	(Multiple $R = 0.999$)	(Multiple $R = 0.972$)
Brand preference (a).....	0.368*	1.108†
Average price (b).....	-0.436*	-0.202
Store coverage (c).....	0.150*	0.609
Stock display (d).....	0.224*	-0.364
Promotional effort (e).....	0.416*	0.067
POP advertising (f).....	-0.242*	-0.207
Advertising expenditure (g).....	0.143*	-0.536*

For cleanser:

$$P_i = 0.368X_1 - 0.436X_2 + 0.150X_3 + 0.224X_4 + 0.416X_5 - 0.242X_6 + 0.143X_7$$

For coffee:

$$P_i = 1.108X_1 - 0.202X_2 + 0.609X_3 - 0.364X_4 + 0.067X_5 - 0.207X_6 - 0.536X_7$$

* Significant at the 5 percent level.

† Significant at the 1 percent level.

termining market shares of brands were price, promotional effort and brand preference. As might be expected, price is negatively related, while promotional effort and brand preference are positively related to market share. One apparent anomaly was that point of purchase advertising was negatively related.

For coffee, the regression model produced a coefficient of multiple correlation of .972, significant at the one percent level of confidence. However, in contrast to the scouring cleanser data, only two of the marketing factors studied, brand preference and advertising expenditure, were found to have significant effects upon the share position of brands of coffee, while store coverage approached significance.

For scouring cleanser it was observed that there were relatively high correlations between the brand preference ratings and several of the variables measuring marketing activity. The question arose—need we consider preference at all in such a demand equation?

The question was answered by dropping preference as an independent variable and noting what happened to the fit of the regression equation. This made little difference: R^2 dropped from .9997 to .9903, a change of less than one percent. The reason for this may be found in the results of the regression of these marketing variables on the preference. Ninety-three percent of the variance in preference for brands of scouring cleanser was accounted for by variance in the six external marketing variables. All of the regression coefficients were significant beyond the five percent level, with those of price, promotional effort and stock display being highest. Differences among these three were not significant.

For coffee, on the other hand, dropping the preference variable reduced R^2 from .9456 to .6063, a change of 35 percentage points. Although the six-variable regression equation without preference for various

brands of coffee still yielded a statistically significant multiple correlation coefficient, it is clear that these customers were more sensitive to the qualities of coffee brands than to the qualities of cleanser brands.

The linear equation based on six marketing variables did a satisfactory job of "explaining" shares of brands of scouring cleanser and coffee. However, there are advantages in reducing the number of variables. Other sets of regression equations were developed, using only the three marketing variables found to have the strongest relation to brand shares.

For scouring cleanser, the three used were price, store coverage and promotional effort. These three variables were quite effective in fitting the data; R^2 dropped from .9903 to .8892, only 10 percent.

Because of the ease with which this three-variable equation could be computed, it was applied to various segments of the total sample. On the basis of information collected during the interview, respondents could be classified in the following ways: by income group; by whether they were exposed to much advertising; and by whether they shopped mostly at chain or independent stores.

Income was determined from 1940 Census rent data and modified by interviewer's evaluation of homes. To be considered as being "exposed to much advertising" they had to be exposed to three advertising vehicles other than radio programs. Type of store usually shopped was determined by questioning.

TABLE 2
REGRESSION COEFFICIENTS BETWEEN CLEANSER BRAND SHARES AND THREE
MARKETING ACTIVITIES: BY STRATA

Group	Regression Coefficients			
	Average Price	Weighted Store Coverage	Promotional Effort	Multiple Correlation Coefficient
Entire sample.....	-0.291†	0.491†	0.603†	0.943†
Chain store.....	0.083	-0.049	1.016†	0.977†
Independents.....	-0.045	0.601*	0.343	0.832*
Advertising prone.....	-0.338†	0.587†	0.530†	0.938†
Non-prone.....	-0.037	0.164	0.831†	0.936†
High income.....	0.044	0.210	0.837†	0.919†
Medium income.....	0.974	0.130	0.867†	0.935†
Low income.....	0.029	0.162	0.809†	0.909*

* Significant at the 5 percent level.

† Significant at the 1 percent level.

In a cross-classification of respondents by income level and stores shopped (see Table 2), it was found that the low income groups patronized independents to a much greater degree than did the two upper income groups. This was largely because few chains have units in the

Negro and low income areas. There was no clear relationship between income and availability to advertising exposure.

Even when the total sample was split up, the regression coefficients for the scouring cleanser data remained significant except for respondents shopping at independent grocery stores. Promotional effort and distribution apparently were more important than price in "explaining" variations in scouring cleanser brand shares for the entire sample as well as for its different segments. However, all three coefficients were still significant beyond the one percent level.

Between the various segments of the sample, some differences in importance of the three variables did emerge. Apparently chain store shoppers were more susceptible to promotional offers and deals for brands of scouring cleanser than housewives who shopped at independent stores; for the latter, availability was the most important factor. Those open to advertising exposure were equally affected by the store distribution of brands and the use of promotional effort.

Price also had a strong effect. Promotional effort was the only variable among the three tested to affect market share among those respondents not available to heavy advertising exposure.

Income seemed to have no effect upon the weights of the variables in the demand equation. This seems plausible since scouring cleanser is relatively cheap. It is interesting to note that disguised price reductions—in terms of special deals or offers—had a much stronger effect upon scouring cleanser brand shares than did actual price differences. This was equally true for all income groups.

A three-variable regression model was also fitted to the coffee data using the variables found to have highest correlation with market shares: price, store coverage and past six months' advertising expenditure.

It was found to yield a statistically significant fit; the coefficient of multiple correlation was significant at the one per cent level. However, the fit of the three-variable equation for coffee was substantially poorer than that for scouring cleanser. There are at least two reasons for this: the greater diversity of marketing patterns among the 21 brands of coffee than among the nine brands of scouring cleanser; and the greater importance of brand quality for coffee than for scouring cleanser.

For the entire sample, only store coverage had a statistically significant relation with market shares of coffee brands. Neither price nor advertising expenditure was found to have a significant effect upon market shares when the other two factors were held constant.

The three-variable model was applied to various segments of the sample and statistically significant fits were obtained among chain store shoppers, those advertising prone and those in the high income group. Probably these three sub-groupings overlapped so that the same respondents showed up under different headings.

Although the data were not statistically significant, an interesting situation held among the "manufacturer's brand" buyers. Among those people who bought a manufacturer's brand (Hills Brothers, Chase & Sanborn, Stewarts, etc.), advertising actually appeared to have a negative relation with sales. Examination of the raw data indicated that, during the period studied, Maxwell House was spending 40 percent of the total advertising volume in Chicago for these six brands, but was receiving only 16 percent of their total sales. In contrast, Manor House was spending only nine per cent of the total advertising volume, but receiving almost 30 percent of total sales.

CONCLUSIONS

The more general model discussed at the beginning of this paper has illustrative value for teaching purposes. It formulates problems of demand in marketing terms by dealing with market share data obtained by differentiated brands whose owners compete with all the tools in their respective arsenals. Students are thus presented with a device for considering the major variables affecting sales.

The demand model easily accommodates the familiar discussion of convenience, shopping and specialty goods. For example, the demand model for convenience goods would likely show store coverage, point of purchase display and promotional effort to be most important in affecting sales. On the other hand, for specialty goods preference would probably be the only variable of major importance.

The model tested by the data presents more of a mixture of values. Such a regression model can approximate the importance of various factors affecting market shares, and the relationships between these factors. Surveys are less helpful on this point since people seldom can evaluate the relative importance of the factors impinging upon their purchase decisions.

The results of regression analysis should be considered as first approximations for several reasons. Foremost is the fact that they show only co-variation, not cause and effect. The regression analysis is useful to point out the factors to be used in experimentation, but should not be considered as a substitute for it.

Findings from the regression model hold only for the range of observations available in the data. Promotional effort was found to have a stronger relationship than price with brand shares of scouring cleanser. But the range of prices was quite narrow, 8.6 to 12.9 cents per can. Whoever breaks through these limits may well find price to have a great effect upon market shares.

Another caution in the use of the regression analysis is that it yields only over-all relationships. In any market, some brands are declining in market shares, others are rising, while still others are merely holding their own. The regression procedure gives coefficients which are actually

averages of the coefficients for the individual brands. These results may not apply to any one brand.

Marketing strategies usually call for manipulating several variables simultaneously. Manufacturers merchandise their coming advertising campaigns to their retailers, who respond by improving stock holdings and displays and by putting up point of purchase advertising sent them. Private brands are usually offered in only a few stores; but in these stores they are usually given the best locations, largest stock displays and massive point of purchase advertising displays. Manufacturers' brands, especially in convenience goods, tend toward 100 percent coverage but with less prominence of display within stores. Regression analysis is not the best way to cope with these different relationships between several independent variables.

Finally, regression analysis is a quantitative procedure and uses essentially quantitative evaluations of data. It is quite likely that many relationships are distorted by the units we use to express these quantities. Advertising is the most important case in point though the problem also arises with premiums and deals. If the effect of advertising were proportional to expenditure on it, then the firm which spent the most for advertising would sell the most product. However, this does not happen (Borden, 1942). Advertising effect depends not only upon magnitude of expenditure but also upon the motivating power of the copy and upon the media used. Failure of advertising expenditure to correlate with sales does not mean that advertising is ineffective, but may mean only that the measuring procedure failed to evaluate properly the strengths of various campaigns.

Implicit in the model presented is the assumption that all variables act instantaneously. This assumption is open to serious doubt. An effort was made to take differing time lags into consideration by considering advertising expenditures for the previous six months, while all other variables were assumed to be acting at the time of the research. This was a guess. It was found that the correlation of brand sales with the previous year's advertising was slightly higher than with the data of the shorter period, but the improvement was not significant. The varying time lag of different variables is certainly one of the most important matters of concern to marketing directors yet little or no research has been devoted to it.

I have said much of the limitations of the regression model but little of its value. I believe it offers real advantages. For relatively little expenditure, a substantial amount of material can be collected and evaluated. The simple correlations between market shares and the independent marketing variables will give a picture of the marketing strategies being used for the brands of a given product class, plus relationships between consumers' appreciation of the qualities of brands and external marketing variables. Finally, the findings of the multiple regression analysis can

be considered a first approximation of the relative importance of these marketing variables—especially if more faith is put in findings of no effect than in findings of much effect.

REFERENCES

- ACTON, F. S. *The Analysis of Straight Line Data*. New York: John Wiley & Sons, Inc., 1959, chap. v.
- ANDERSON, R. L. and BANCROFT, T. A. *Statistical Theory in Research*. New York: McGraw-Hill Book Co., Inc., 1951.
- BANKS, SEYMOUR. "The Relationships between Preference and Purchase of Brands," *Journal of Marketing*, Vol. XV, No. 2 (October, 1950), pp. 145-57.
- BARTLETT, M. S. "Fitting a Straight Line When Both Variables Are Subject to Error," *Biometrics*, Vol. V (1949), pp. 207-12.
- BORDEN, N. H. *The Economic Effects of Advertising*. Chicago: Richard D. Irwin, Inc., 1942, chap. viii.
- EZEKIEL, M. *Methods of Correlation Analysis*. 2d ed. New York: John Wiley & Sons, Inc., 1941, chap. xxi.

Long-Term Trends in Food Consumption: A Multi-Country Study*

.....

L. JURÉEN†

1. Food Demand at Rising Income Level¹

IN ORDER TO SHOW IN GENERAL OUTLINE THE EFFECT OF RISING INCOME (or varying price levels) on food demand it is convenient to aggregate all food items into a few large groups. The simplest procedure is to split up the total calorie intake into two groups only, one containing cheap food items or "necessities" and the other the remaining expensive foods or items of a more "luxury" character. From the agricultural standpoint it is natural to deal with animal foods (except nonagricultural products such as fish and marine oils) as one group. Then the second group becomes rather heterogenous, including not only predominantly cheap products such as cereals and potatoes, but also the somewhat more expensive sub-groups, sugar and vegetable oils and fats and, furthermore, fish, fruit and vegetables, which are luxuries if the calorie content only is taken into account. Therefore, the demand for cereals as well as fruit and vegetables is treated separately below and some attention is also paid to the consumption of other main items.

Figure 1 is a scatter-diagram, showing pre-war calorie intake per capita per day within the two groups of food items—animal foods and residual items—in sixteen European countries² and the United States as plotted against average real income per head of the population. The basic consumption data are taken from the FAO food balance sheets, and data

* Reprinted from *Econometrica*, Vol. XXIV, No. 1 (January, 1956), pp. 1-21.

† Stockholm, Sweden.

¹ This report is based upon results obtained by the author in a research project at the Economic Commission for Europe (ECE), United Nations, Geneva. Some preliminary results are published in a statement prepared by ECE and FAO (*European Agriculture—A Statement of Problems*, Geneva, 1954, Chapter 2 and Appendix A). The results have also been used in a report to the World Population Conference in Rome in September, 1954 (L. Juréen and H. Wold: *The Regional Forecasting of Food Demand*).

² That is, all countries for which statistical data have been available.

representing real income are from the *Economic Survey* in 1949.³ Simple hyperbolic demand curves are fitted to the data concerning total calorie intake and animal products. The demand function for "vegetable products" (i.e., all residual items) is represented by the differences between the hyperbolic curves.⁴ The demand functions $q(r)$ and the income elasticities $E(r)$ obtained are as follows where r is real income, a refers to the aggregate of animal foods, v to the aggregate of all other foods, and t to total calorie intake:

$$\begin{aligned} q_a(r) &= \frac{1690r^*}{r + 134} & E_a(r) &= \frac{134}{r + 134} \\ q_v(r) &= q_t(r) - q_a(r) & E_v(r) &= \frac{E_t(r) \cdot q_t(r) - E_a(r) \cdot q_a(r)}{q_t r - q_a r} \\ q_t(r) &= \frac{3308r}{r + 13} & E_t(r) &= \frac{13}{r + 13} \end{aligned}$$

Despite differences between countries in respect to geographical position, traditional consumption habits, and price relations (and, in addition, despite fairly large errors surrounding the data used) a certain, though not linear, relationship is apparent between food consumption and real income. Certainly the discrepancies between the demand curves and the original data are large in many cases, but most of them can be explained. For instance, the calorie intake in the Mediterranean region would still be lower than in western and eastern Europe even if the levels of living were the same, the reason being that of different climatic conditions. This seems also to be the main reason why total calorie intake as well as consumption of animal foods is lower in Italy than is indicated by the curves, which already have rather low values at the Italian in-

³ *Economic Survey of Europe in 1949*. Economic Commission for Europe. Geneva, 1950. Table IV, total commodities available per head; prewar values expressed in U.S. dollars.

⁴ Linear curves (or linear, when transformed to logarithms) cannot be used as approximations to the data because of the great range of income variation. The hyperbolic function used is one of the so-called "Törnqvist demand functions." The fitting is performed by the method of least squares.

* Editor's note: These demand functions can be changed into the following form by simple algebraic manipulation:

$$\frac{1}{q_a(r)} = \frac{1}{1690} + \frac{134}{r}$$

The regression would take $1/q_a(r)$ as the dependent variable and $1/r$ as the independent variable.

The income elasticity is defined as:

$$E_a(r) = \frac{dq_a(r)}{dr} \cdot \left\{ \frac{r}{q_a(r)} \right\}$$

where $dq_a(r)/dr$ is the derivative of $q_a(r)$ with respect to r . The value of $E_a(r)$ was derived from the expression for $q_a(r)$ by applying this formula.

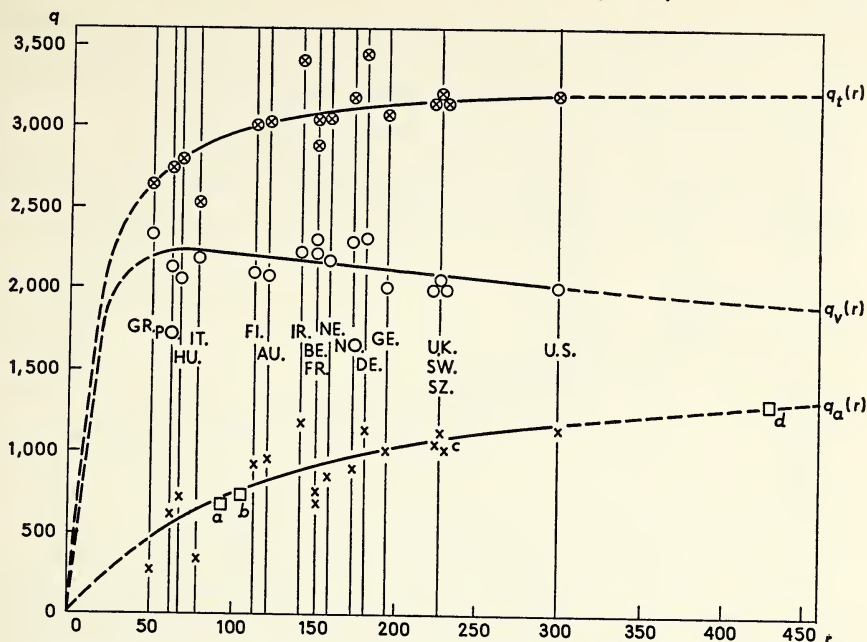


FIGURE 1. Countries are identified as follows: Gr = Greece, Po = Poland, Hu = Hungary, It = Italy, Fi = Finland, Au = Austria, Ir = Ireland, Be = Belgium, Fr = France, Ne = Netherlands, No = Norway, De = Denmark, Ge = Germany, UK = United Kingdom, Sw = Sweden, Sz = Switzerland, and US = United States.

Special points in the diagram are labeled as follows: a = farm and forestry workers; b = small farmers; c = industrial workers and low grade employees; and d = middle class.

come level. The very high consumption of animal foods in Ireland and Denmark is also explained—at least to some extent—by the climatic factor. Furthermore, the latter countries are traditional exporters of animal products, which means that prices show a tendency to be relatively low and, accordingly, consumption rather high. In this qualitative manner differences can be explained country by country. The disturbing factors referred to, however, are not introduced in relations (1). Obviously, if they were in the contrary case they might not give a random effect at all but instead exert a systematic influence upon the slope of the demand curves. This problem could not be solved by introducing additional variables in (1), especially as the quantitative effect of varying climatic conditions on food consumption is not very well known.

Because relative food prices have shown extraordinary changes since the prewar period,⁵ a study of the post-war pattern of food consumption at different income levels (price changes being disregarded) would give some information about the stability of the curves. Using post-war data for all countries listed in Figure 1 except Hungary (data not available),

⁵ *European Agriculture—A Statement of Problems*, ECE and FAO, Geneva, 1954, pp. 28–30.

we find that the pre-war pattern repeats itself fairly well; this is shown in Figures 2 and 3. In spite of the abnormal price structure in 1949-51, the slope of the curves fitted to cereals and to animal foods is almost unchanged. (For sources and methods used, see the Appendix.)

Of course, errors caused by factors which are rather stable in time, country by country (e.g., climatic conditions, etc.), cannot be shown by comparisons of the kind used for Figures 2 and 3. The question re-

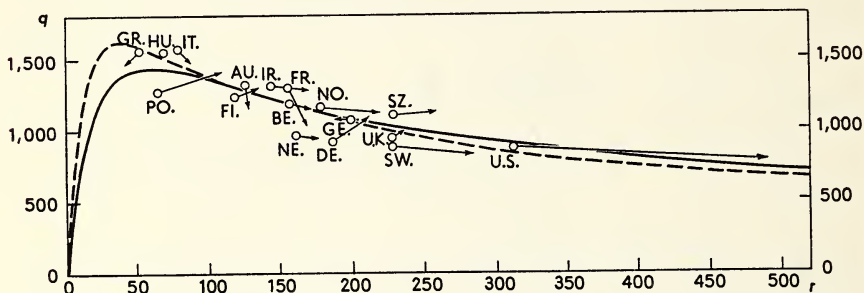


FIGURE 2. Countries are identified as in Figure 1. q = consumption in calories per capita per day; r = real income per capita in prewar U.S. dollars. Observations indicated by small circles refer to prewar data, as does the broken regression curve; observations indicated by the tips of arrows refer to postwar data, as does the solid regression curve.

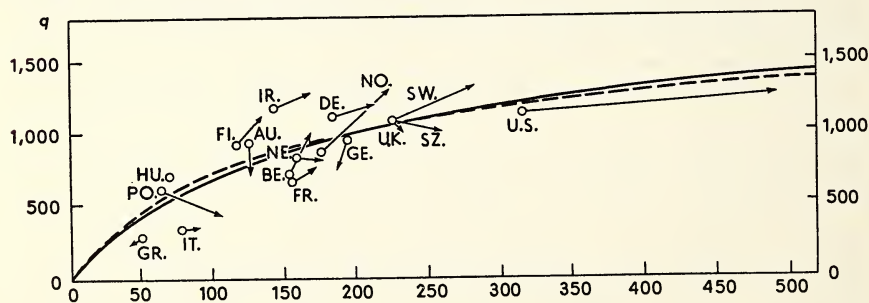


FIGURE 3. Units and notation are as in Figure 2.

mains, therefore, whether the inter-country comparisons—in spite of the errors just mentioned—are capable of being used to furnish quantitative information, e.g., about (a) probable differences in food consumption by classes of the population, and (b) the probable long-term consumption development within *one* country. These questions must be considered before the results are used as indicators.

Data shedding light on (a) and (b) are far from satisfactory. Some illustrations can be obtained, however, by use of statistics referring to Sweden and the United States. Beginning with the question of consumption by classes of the population, Table 1 shows income elasticities for animal food in Sweden compared with results obtained from Figure 1.

The dependence of elasticity upon income is seen clearly in the

TABLE 1
INCOME ELASTICITIES FOR ANIMAL FOOD: SWEDISH DATA, 1933

Social Class	Sweden: Real Income, kr. per Family of 3.3 Consumption Units	Corre- sponding Income Level in Figure 1	Income Elasticities in Sweden		
			Actual Estimates	Smoothed Values	Obtained from Figure 1
Farm and forestry workers . . .	1,087	93	0.73	0.62	0.59
Small farmers	1,245	107	0.50	0.59	0.55
Industrial workers and low- grade employees	2,666	227	0.39	0.41	0.37
Middle class	5,049	431	0.27	0.26	0.24

Sources and methods: See Appendix.

Swedish data. In classes with low incomes, such as farm workers and small farmers, the elasticity for animal food is found to be the highest. Industrial workers and low-grade employees have considerably higher incomes, and consumption is here less sensitive to variation in the income level. In middle-class families with relatively high incomes consumption is least influenced by differences in income.

The question is, however, whether the decrease in the income elasticity is mainly due to the rising income level or whether the tendency is to a considerable degree due to the influence of other factors. The Swedish investigation referred to comes to the conclusion (supported by empirical findings not quoted here) that the income level is the factor of primary importance behind the differences in the income elasticities for animal food (as well as for other food items), and that the influence of social factors on consumer habits is small or negligible.

The demand function values corresponding to the smoothed elasticity values are plotted in Figure 1. The fairly close agreement between the elasticities recorded above causes the plotted values to lie very near the demand curve fitted to the FAO data. The combined result suggests the conclusion that differences in the average animal consumption levels are, not only between classes of the population but also between countries, primarily determined by the average income level, while differences in geographical position, consumption habits, and income distribution seem to have a relatively small influence on consumption. In any case the disturbing influence on the variable elasticity values that can be derived from Figure 1 seems to be very small, and this is of crucial importance for the practical use of the results.⁶

⁶ Of course, this argument is valid only for the entire aggregate of animal products but not necessarily for specified items. It may also be pointed out that the conclusions refer to Europe and the United States only. Certainly they cannot be extended so as to include also the poorer overseas countries where the influence of geographical position on food production cannot be disregarded and, consequently, the consumption habits are "traditional" in the sense that the very low import of foodstuffs does not in any appreciable degree permit variation from what is determined by the agricultural conditions.

Coming to the question of the development of consumption through time within countries, it is convenient to show the long-term trends in countries where the nutritional standards are high *today*. Table 2 shows data for Sweden and Table 3 similar data for the United States. True, the level of living has been for a long time and still is perceptibly higher in the latter country, but differences with regard to nutrition are now very small.⁷

The period reviewed in Table 2 has been marked in Sweden—as in

TABLE 2
PER CAPITA CONSUMPTION OF FOOD IN SWEDEN SINCE 1876 IN CALORIES PER DAY

	1876- 85	1886- 95	1896- 1905	1906- 13*	1920- 39	1930- 39	1940- 47	1952
Meat, bacon, eggs....	190	220	300	320	370	420	380	460
Margarine.....		10	40	60	120	190	120	280
Milk, cream, cheese, butter.....	360	380	490	610	700	760	780	800
<i>Total of above..</i>	550	610	830	990	1190	1370	1280	1540
Vegetable products†..	1730	1820	1940	1990	1860	1730	1690	1630
<i>All food‡.....</i>	2280	2340	2770	2980	3050	3100	2970	3170

* For vegetable products, 1906-15.

† Flour and grits of wheat, rye, oats and barley; peas and beans; sugar and potatoes.

‡ All important foods are included, with the exception of fruit, green vegetables, and fish, which correspond at the present time to about 5 percent of the calorie intake of all foods.

Sources: See Appendix.

other industrial countries—by a rapid increase in the levels of living. As far as Sweden is concerned, national income has increased since 1870 at an average rate of about 2 percent per head per year. (This means a doubling of national income per head from one generation to the next.) The consumption data referring to 1876-85 correspond roughly to the income level of 70 in Figure 1.⁸

As indicated in the table, the per capita consumption of cheaper vegetable foods reached its peak before World War I, since which time a steady decline has occurred. (The sub-group cereals showed the same development.) The consumption of animal products climbed from one-fourth of the total calorie intake in 1880 to about one-third prior to World War I and is now approaching half of the total consumption. Consumption has, however, increased at a declining rate both for milk and dairy products as a group and for the aggregate of animal products. The

⁷ The levels of calorie and protein intake are about the same. Fat consumption is somewhat higher in the United States—perhaps too high from a nutritional standpoint. The consumption of carbohydrates (in grams) is on the same very low level in both countries. See also Figure 1.

⁸ This value makes it possible to plot the data from the table somewhere around the right place on the income scale in the figure. However, it may be observed that the aggregation of food items is not exactly the same in the table as in the figure.

TABLE 3
PER CAPITA CONSUMPTION OF FOOD IN THE UNITED STATES SINCE 1909
(1909-16 = 100)

	1909-16	1923-29	1935-39	1951/52
Flour, groats, and bread.....	100	84	73	61
Sugar.....	100	127	118	112
Potatoes.....	100	86	77	55
Fruits and vegetables.....	100	115	133	171
Meat and fish.....	100	95	89	104
Eggs.....	100	109	97	135
Oils and fats, including butter.....	100	105	99	99
Milk and dairy products, excluding butter.....	100	122	123	153

Source: See Appendix.

time is still far off, however, when per capita animal consumption will reach its maximum level.

It is interesting to compare the figures for Sweden in 1906-29 and for the United States in 1909-29, bearing in mind the higher level of living in the latter country. The total per capita calorie intake of cereals, sugar, and potatoes dropped in both countries during a period of roughly 15 years by 5 to 6 percent. In both countries animal consumption increased, but the rate of increase was considerably slower in the United States, where for the whole animal group (including vegetable oils and fats) it was about 4 percent as against 20 percent in Sweden, and for the sub-group of milk and dairy products, vegetable oils, fats and margarine it was about 11 percent as against 22 percent. It may be added that in the inter-war period 1920-39 the annual increase in animal consumption, like that in the consumption of milk and dairy products, was still considerably higher in Sweden than it was in the United States during the period 1909 to 1929.

Between 1923-29 and 1935-39 animal consumption (including vegetable oils and fats) dropped by 5 percent in the United States. Afterwards the consumption of eggs and milk increased substantially. On the other hand, the data given here indicate that a stagnation has occurred in the United States in the consumption of meat and fish and of oils and fats.

Long-term data of the kind now recorded can be compared with the demand curves in Figure 1. Without going into details in this matter, it may be said that, insofar as the United States and Sweden are concerned, the course of the long-term trends is in fairly close agreement with the shape of the curves in the figure. In other words, judging from the time series available, multi-country curves of the type chosen can be used as a framework for outlining what happens with food consumption in the long run in a country when average income rises. They afford a possibility of showing roughly—but in concrete terms—the probable trends

of main groups of food items at a certain income level. As empirical support for this somewhat tentative conclusion, it may be sufficient here to give only one more example. Table 4 shows elasticities, based on yearly

TABLE 4
INCOME ELASTICITIES FOR AGGREGATED FOOD ITEMS AT A HIGH PREWAR INCOME LEVEL*

<i>Income Elasticities Derived from Swedish Time Series</i>					
<i>Aggregate</i>	<i>1921-39</i>		<i>1923-38</i>		<i>Multicountry Curves</i>
Animal foods.....	0.36	(0.02)	0.35	(0.03)	0.37
All other food items.....	-0.12	(0.06)	-0.26	(0.05)	-0.11
of which cereals represent.....	-0.56	(0.09)			-0.37
Total food in calories.....			0.07	(0.03)	0.05
Total food volume in constant prices..	0.28	(0.02)	0.23	(0.04)	0.23

* 227 U.S. dollars per head (Swedish average level).
Sources and methods: See Appendix.

time-series data, for Sweden in the inter-war period (standard errors given in brackets) and the corresponding values obtained from the multicountry curves at the Swedish pre-war average income level. On the whole the agreement between the results is satisfactory.

Direct estimates which have been made of income elasticities in the United Kingdom and the United States, however, agree less closely, as the data below show in comparison with elasticities obtained from Figure 1.⁹

	Income Elasticity	
United Kingdom:		
From Figure 1 (1934-38).....	0.23	
Time series, 1920-38 (Stone).....	-0.02	
Family budgets (Stone).....	0.53	
United States:		
From Figure 1 (1935-38).....	0.20	
Time series (sometimes in combination with family budget results)		
1913-41 (Tobin).....	0.27	
1929-41 (Girshick and Haavelmo).....	0.25	(plus 0.05, which relates to last year's income)
1929-41 (Stone A).....	0.59	
(Stone B).....	0.53	
(Stone C).....	0.83	
1913-41 (Tobin-Stone).....	0.61	

The discrepancy in the case of the United Kingdom is particularly striking; it can only to some minor extent be explained by the fact that

⁹ The sources are: R. Stone, *The Demand for Food in the United Kingdom*, Cambridge, 1950; J. Tobin, "A Statistical Demand Function for Food in the U.S.A.," *Journal of the Royal Statistical Society*, Pt. II, 1950 and discussion by R. Stone in the same issue; and M. A. Girshick and T. Haavelmo, "Statistical Analysis of the Demand for Food: Examples of Simultaneous Estimation of Structural Equations," *Econometrica*, Vol. XV (1947), pp. 79-110.

TABLE 5

PREWAR AND POSTWAR INCOME ELASTICITIES AT VARYING INCOME LEVELS

Income Level per Head, in U.S. Dollars	Income Elasticities for Aggregated Food Items											
	All Food Items Other than Animal Foods				Cereals				Total Food in Calories			
	Animal Foods		Postwar		Prewar		Postwar		Prewar		Postwar	
	Prewar	Alt. I.	Alt. II.	Prewar	Alt. I.	Alt. II.	Prewar	Alt. I.	Alt. II.	Prewar	Alt. I.	Alt. II.
25	80	(0.84)	(0.82)	(0.87)	(0.27)	(0.31)	(0.38)	(0.15)	(0.21)	(0.27)	(0.34)	(0.38)
35	110	(0.79)	0.76	0.83	(0.18)	0.22	0.28	(0.03)	0.08	0.15	(0.27)	0.31
50	160	0.73	0.69	0.77	0.09	0.12	0.18	-0.10	-0.05	0.01	0.20	0.24
75	240	0.64	0.60	0.68	0.01	0.04	0.07	-0.23	-0.18	-0.14	0.15	0.17
100	320	0.57	0.53	0.62	-0.04	-0.01	0.01	-0.31	-0.25	-0.23	0.11	0.14
125	400	0.52	0.47	0.57	-0.07	-0.04	-0.03	-0.37	-0.29	-0.29	0.09	0.11
150	480	0.47	0.43	0.52	-0.08	-0.06	-0.06	-0.40	-0.32	-0.34	0.08	0.09
200	640	0.40	0.36	0.45	-0.10	-0.08	-0.09	-0.45	-0.34	-0.39	0.06	0.07
250	800	0.35	0.31	0.39	-0.11	-0.08	-0.11	-0.47	-0.35	-0.42	0.05	0.06
300	960	0.31	0.27	0.35	-0.12	-0.09	-0.12	-0.48	-0.35	-0.44	0.04	0.05
400	1,280	(0.25)	0.22	0.29	(-0.11)	-0.08	-0.12	(-0.49)	-0.33	-0.45	(0.03)	0.04
600	1,920	(0.18)	(0.16)	(0.22)	(-0.10)	(-0.07)	(-0.12)	(-0.46)	(-0.29)	(-0.43)	(0.02)	(0.03)
											(0.50)	(0.55)
											(0.47)	0.49
											0.42	0.44
											0.37	0.38
											0.34	0.34
											0.31	0.36
											0.29	0.28
											0.25	0.24
											0.22	0.21
											0.20	0.18
											0.17	0.15
											(0.12)	(0.11)

Note. The italicized values refer to the average income level for Europe as a whole. Elasticities referring to income levels outside the interval covered by consumption data are put in parentheses. In general, prewar data refer to 1934-38; postwar data refer to 1949/50-1950/51.

Sources and Methods: See Appendix.

budget data in this case should give higher results than time-series data.¹⁰ As to the various estimates for the United States, it may be pointed out that the last four estimates quoted above are based on consumption data including changes in the whole cost of processing and are therefore not comparable with elasticities obtained from Figure 1. (These costs are deliberately left out here because a rising amount of processing does not mean an increase of quantities sold by farmers.) The first two estimates for the United States are less influenced by the amount of processing and they agree quite well with the result from Figure 1, which, in fact, should have a somewhat higher value than 0.20 when taken as an average for the periods of the time series chosen.

Elasticities at varying income levels are recorded in Table 5. These are values obtained from the multi-country curves. As will be seen from the table the demand for animal foods grows steadily higher at higher income levels but the elasticity is much lower at higher levels than at the lower ones. The possible improvement in animal consumption is, however, still significant, the average income elasticity for Europe as a whole having a value around 0.5. In the poorer regions—say, countries with an income level lower than the European average, i.e., countries having almost 60 percent of the total European population within their boundaries—the average elasticities probably range from 0.5 to 0.8, and in the other countries—roughly the northwestern European region—from 0.35 to 0.5.

As for the heterogeneous group “all other food items,” the aggregated demand is almost inelastic except in the very poorest countries. The main subgroup included is cereals,¹¹ and special attention has to be drawn to those items. Here consumption shows, broadly speaking, three phases: (a) rising demand in countries with very low levels of living; (b) constant demand in countries with still rather low levels; and (c) falling demand in countries with medium or high levels. As far as could be judged from comparisons among countries, southern Europe, except Italy, is in the first phase, while northwestern Europe is in the third. Italy seems to be the most important country in the middle

¹⁰ It may be argued that—as a result of the continual introduction of novel commodities into the market—the income elasticities of family budget data on the whole tend to be smaller than the income elasticities that refer to market statistics (see H. Wold in association with L. Jürén: *Demand Analysis* [Stockholm, 1952; New York, 1953], chap. 14.7). In this case a rather strong reverse influence on the average elasticity values is apparent, however, and is caused by the fact that certain elastic items (processed foods, consumption in restaurants, and so on) are included in the set of budget data but not in the market statistics used.

¹¹ Sugar is another important item in the residual group. This commodity, although a rather cheap one, belongs to the food items that show increasing demand at rising standards of living (see for instance *Sugar*, FAO Commodity Series, September, 1952). Compared with prewar data, the consumption is now roughly on the same level in countries with high or medium real incomes per head, but it is substantially higher in most of the poorer European countries.

phase. In the eastern European countries the demand for cereals seems to be either growing or constant when income shows an average increase. Thus—assuming constant real food prices—the peak in cereal consumption is reached at a rather low income level. This statement is also supported by Table 3, showing that cereal consumption in the United States has steadily decreased since 1906–16; the peak was probably reached already in the last decades of the nineteenth century. As to Sweden, it may be recalled that the peak occurred in 1906–15; at that time the real income level, corresponding to the income scale in Table 5, can be fixed somewhere between 100 and the pre-war European average of 125.

The elasticities for total food demand in calories confirms the well-known fact that the calorie intake has almost reached a maximum level in the rich countries, while a certain increase is to be expected in the poorest countries if economic conditions are improved. From the agricultural standpoint the figures in the last line in Table 5 are more interesting because they give estimates of the probable increase in the scope of the market at rising standards of living. In the northwestern European countries total food volume is now fairly stable, income elasticity being 0.2 or 0.3. On the other hand, the elasticity is substantially higher in the other countries, perhaps 0.4 or 0.5, or even higher in the poorest regions.

It may be pointed out that the elasticity values referred to above are valid, strictly speaking, only when the income distribution remains unchanged. If an average income rise in a country were mainly the result of an improvement in the level of the lowest income groups, food demand would increase more rapidly than shown above. In countries with high average levels this tendency is small or negligible, but in poorer countries it must be taken into account. In Table 4, for instance, the elasticities referring to Sweden are possibly influenced by the actual income-levelling that took place during the inter-war period; however, the values are not significantly higher than those obtained from the multi-country curves. On the other hand, taking an extreme example, in countries with an income distribution like Italy's, the improvement in the level of the lowest income groups alone means that the average elasticity of total food volume could be up to 50 percent higher than in the case of an unchanged income distribution. The development of French consumption since 1934–38 may be taken as another example. The income-levelling effect caused by high family allowances introduced in recent years seems to be one of the reasons for the improvement in the levels of food-consumption since the war, as is shown by French national statistics.

2. The Influence of Changes in Food Prices

The relationship between demand and food prices is briefly analysed below by using the results of studies of specified countries as well as inter-country comparisons. The basic material used in the inter-country comparisons is recorded in Table 6.

TABLE 6
FOOD CONSUMPTION DISTRIBUTION AND CALORIE PRICE RATIOS IN 1950/51

Country	Calorie Intake				Retail Price Values				Calorie Price Ratios		
	Animal Prod- ucts† q_a	Cereals and Potatoes q_c	Fruit and Vege- tables q_f	Total N.W. Europe (average = 100)	Animal Prod- ucts†	Cereals and Potatoes	Fruit and Vege- tables	Total	p_a/p_c	p_f/p_c	p_f/p_a
Austria.....	28	67	5	100	53	29	18	100	4.5	8.1	1.8
Denmark.....	42	53	5	100	67	17	16	100	5.0	9.2	1.8
France.....	32	62	6	100	66	17	17	100	7.5	10.0	1.3
Western Germany.....	31	64	5	100	61	21	18	100	5.9	11.7	2.0
Netherlands.....	33	61	6	100	62	20	18	100	5.9	9.6	1.6
Norway.....	38	59	3	100	68	19	13	100	5.7	16.3	2.9
Sweden.....	50	46	4	100	66	19	15	100	3.2	9.2	2.8
Switzerland.....	39	53	8	100	69	14	17	100	6.6	8.5	1.3
United Kingdom.....	43	52	5	100	69	15	16	100	5.7	11.2	2.0
North-West Europe, weighted average*	36	59	5	100	66	18	16	100	...	...	...
Italy.....	14	79	7	100	51	34	15	100	8.1	5.1	0.6

* Total population data have been applied as weights to the absolute figures.

† Meat, eggs, milk, cheese, and butter.

Source: See Appendix.

Food items are aggregated into three groups, namely: (a) animal foods, (b) cereals and potatoes, and (c) fruit and vegetables.¹² Items in aggregate (b) can be regarded as necessities, the consumption of which is well satisfied even at rather low income levels in the area concerned (northwestern Europe). Items in (a) and (c), on the other hand, are more of a luxury character, consumption being low at low income levels but showing a somewhat rapid growth when income increases. Elasticity calculations, made in various countries and based on family budget data, have given coincidental results insofar as the income elasticity of aggregate (b) is very low or nil (or negative at higher income levels), while the elasticities of aggregates (a) and (c) are higher, though the demand is still generally under-elastic (income elasticities less than unity).

When the possibility of substitution between items (or aggregates) is small, the price elasticities (taken with opposite signs) will roughly show the same pattern that the income elasticities do. The substitution problem cannot, however, be neglected in dealing with food demand. Taking income elasticities for the three aggregates (e.g., obtained from family budget data) as estimates of price elasticities means, in fact, that we are using approximate sums of price and cross elasticities. This is a correct (though sometimes crude) method of estimating price elasticities when the average price changes for the three aggregates are supposed to be parallel. But it is incorrect if the price changes are isolated or of different strength and the cross elasticities have values distinctly different from zero. Thus, changes in food demand cannot be fairly analysed without taking the cross elasticities into account.

Coming to the question of the type of demand function to be chosen in this case, we may start by assuming that the average income level remains stable, i.e., that income elasticities can be left out of the picture; and that the cross elasticities for total food demand with respect to the prices of items other than food can be neglected. Then the demand for the aggregates of food items can be made to depend upon three explanatory variables, the average price of animal foods, cereals and potatoes, and fruits and vegetables. For total food demand the average price of all food items, p_t , can be used as a single explanatory variable. Choosing demand functions with constant elasticities, we then have the following relations:

$$\begin{aligned} q_a &= k_a p_a^{e(a,a)} p_c^{e(a,c)} p_f^{e(a,f)} \\ q_c &= k_c p_a^{e(c,a)} p_c^{e(c,c)} p_f^{e(c,f)} \\ q_f &= k_f p_a^{e(f,a)} p_c^{e(f,c)} p_f^{e(f,f)} \\ q_t &= k_t p_t^{e(t,t)} \end{aligned} \quad (2)$$

where q is a demanded quantity, p an average price, e a price or cross elasticity, k a constant and a, c, f, t refer to the aggregates.

¹² The group "animal foods" includes meat, eggs, milk, cheese, and butter. Fish, not being an agricultural product, is excluded. Other main items excluded are vegetable oils and fats and sugar, the reason being that shifts in demand are not closely connected with changes in prices and incomes.

3. An Initial Attempt at Estimating the Elasticities

Up to now comprehensive studies made on demand elasticities for aggregates of food items have not often been published. A recent study¹³ based on Swedish data, however, gives, *inter alia*, the results shown in Table 7¹⁴ (standard errors are given in brackets).

TABLE 7
DEMAND ELASTICITIES FOR FOODS IN SWEDEN

		1921-39		1923-38	
Animal foods:*	$e(a,a)$	-0.37	(0.11)	-0.49	(0.11)
	$e(a,c)$	0.01	(0.10)	0.04	(0.10)
Cereals, potatoes and sugar.*	$e(c,c)$	-0.05	(0.17)	-0.02	(0.19)
	$e(c,a)$	0.36	(0.19)	0.40	(0.20)
All agricultural products.*	$e(t,t)$	0.01	(0.08)		
Total food demand I*	$e(t,t)$	-0.08	(0.09)		
Total food demand II†	$e(t,t)$	-0.21	(0.05)	-0.23‡	(0.06)

* Calorie intake.

† Measured by quantity index (quantity in constant prices).

‡ 1926-37.

In words, the results indicate that the price elasticity of the demand for animal foods has a significant value about -0.4 or -0.5 , while the cross elasticity is nearly zero. On the other hand, the low price elasticity for cereals, potatoes, and sugar shows that demand has no significant dependence on the price of this food group, while the cross elasticity has a value around 0.40 , which cannot be neglected. Total calorie intake is almost entirely independent of price changes, but this is not the case when total

TABLE 8
DEMAND ELASTICITIES FOR FOODS IN THE UNITED KINGDOM

DEMAND ELASTICITIES FOR FOODS IN THE UNITED KINGDOM					
		Income Elasticities			
		Price Elasticity	United Kingdom	Sweden	
				Workers	Middle Class
Apples	-0.81 (0.23)	1.33	0.84	0.70	
Oranges	-0.93 (0.28)	0.92			
Bananas	-0.51 (0.21)	0.95			
Other fresh fruit	-0.93 (0.25)	1.62	0.46*	0.66*	
Fresh green vegetables and legumes	-0.31 (0.27)	0.93			
Root vegetables, tomatoes, etc.	-0.47 (0.20)	0.85			

* Including potatoes.

demand is defined as a volume measured in constant prices, price elasticity then being around -0.2 .

¹³ L. Jurén, *The Agricultural Production and Food Consumption in Sweden* (in press 1955), S.O.U.

¹⁴ See also Wold-Jurén, *Demand Analysis*, Table 17.6.2.

The Swedish study gives no information on price and cross elasticities for fruit and vegetables, the reason being the lack of reliable market statistics. Some results are, however, given in a work by R. Stone concerning food demand in the United Kingdom¹⁵ and are summarized in Table 8 (standard errors are given in brackets). This table is supplemented with income elasticities for the United Kingdom (Stone) and Sweden (Wold-Jurén).

In the United Kingdom the price-elasticity estimates are rather small compared with the income elasticities obtained. Perhaps part of the discrepancy is due to errors in the market data used. This is indicated by the somewhat high values of the standard errors (the presence of errors in the original data means that elasticities obtained by regression analysis are underestimated). It seems reasonable to use a value of about -1.0 to represent the average price elasticity for fruit and a value of about -0.5 for vegetables. The weighted average for fruit and vegetables is then about -0.80 .¹⁶

The above results have been summarised as follows:

$$\begin{aligned} e(a, a) &= -0.45 & e(a, c) &= 0.05 \\ e(c, c) &= -0.05 & e(c, a) &= 0.40 \\ e(f, f) &= -0.80 \\ e(t, t) &= -0.20 \end{aligned}$$

Inserting these results in (2) we get:

$$\begin{aligned} q_a &= k_a p_a^{-0.45} p_c^{0.05} p_f^{e(a,f)} \\ q_c &= k_c p_a^{0.40} p_c^{-0.05} p_f^{e(c,f)} \\ q_f &= k_f p_a^{e(f,a)} p_c^{e(f,c)} p_f^{-0.80} \\ q_t &= k_t p_t^{-0.20} \end{aligned} \quad (3)$$

Before dealing with the remaining cross elasticities in (3) that are not yet known, we must determine whether the combined empirical findings referring merely to Sweden and the United Kingdom can be used as estimates for the northwestern European area as a whole without any modification.

4. An Experiment Based on National Statistics from 10 Countries

Tables 11 and 12 in the appendix show consumption data and price ratios for ten European countries. A direct comparison between these data is not feasible because the average income level varies from country to country. The number of countries concerned is too small to permit the use of income as an extra variable in the analysis. Instead, the consumption

¹⁵ R. Stone, *The Demand for Food in the United Kingdom* (Cambridge, Eng., 1960).

¹⁶ The consumption values of fruit and vegetables respectively (averages for northwestern Europe) have been used as weights.

data have been adjusted in advance in order to show the probable distribution of consumption at a constant income level. The result is shown in Table 9 (for statistical methods used see the Appendix).

TABLE 9
ESTIMATED CONSUMPTION DISTRIBUTION AT CONSTANT INCOME LEVEL

ESTIMATED CONSUMPTION DATA							
Country	Actual Consumption			Estimated Consumption at Income Level = 100			
	Animal Products	Cereals and Potatoes	Fruits and Vegetables	Actual Income Level	Animal Prod- ucts	Cereals and Vegetables	Fruit and Vegetables
Austria	28	67	5	60	34	59	7
Denmark	42	53	5	100	42	53	5
France	32	62	6	70	36	56	8
Western Germany	31	64	5	70	35	59	6
Italy	14	79	7	40	21	68	11
Netherlands	33	61	6	70	37	55	8
Norway	38	59	3	90	39	58	3
Sweden	50	46	4	115	47	49	4
Switzerland	39	53	8	100	39	53	8
United Kingdom	43	52	5	100	43	52	5

The relationship between estimated consumption and price ratios has been analysed by means of ordinary regression techniques, using the following formula:

$$q_a = k'_a(p_a/p_c)^{e(a/c)}(p_a/p_f)^{e(a/f)}$$

and similar formulas for q_c and q_f .

The results obtained are shown below:¹⁷

$$q_a = k'_a(p_a/p_c)^{-0.03}(p_a/p_f)^{-0.39} \quad R = 0.84 \quad (4.1)$$

$$q_c = k'_c(p_c/p_a)^{-0.19}(p_c/p_f)^{0.09} \quad R = 0.68 \quad (4.2)$$

$$q_f = k'_f(p_f/p_a)^{-0.69}(p_f/p_c)^{-0.24} \quad R = 0.90 \quad (4.3)$$

To get results comparable with (3) we must introduce the additional variable p_a in (4.1), p_c in (4.2) and p_f in (4.3). In practice this is not possible for two reasons: (a) comparisons between absolute prices country by country are doubtful, and (b) the relationship between the explanatory variables generally becomes almost functional, i.e., the cross elasticity values become indeterminable. In theory, however, we get

$$q_a = k''_a(p_a/p_c)^{-0.03+\alpha}(p_a/p_f)^{-0.39+\beta}p_a^{-\gamma} \quad (5)$$

and similar formulas for q_c and q_f .

Relation (5) can be written

$$q_a = k''_a p_a^{-0.42+\alpha+\beta-\gamma} p_c^{0.09-\alpha} p_f^{0.35-\beta} \quad (6)$$

When the relationship between the explanatory variables tends towards

¹⁷ Standard errors of elasticities are, respectively, for (4.1), ± 0.25 and ± 0.15 ; for (4.2) ± 0.10 and ± 0.09 ; and for (4.3) ± 0.25 and ± 0.39 .

a functional one, the value of $(\alpha + \beta - \gamma)$ is almost zero. Then -0.42 can be used as an estimate of the price elasticity $e(a, a)$. Estimates of the elasticities $e(c, c)$ and $e(f, f)$ are obtained in a similar way. The results may be compared with the corresponding values for Sweden and the United Kingdom arrived at earlier:

	Sweden and the United Kingdom	10 European Countries
$e(a, a)$	-0.45	-0.42
$e(c, c)$	-0.05	-0.10
$e(f, f)$	-0.80	-0.93

The almost coincidental results are obtained by using quite different sets of data. This fact may possibly be taken as a sign that the price elasticities found have fairly reliable values, though the standard errors are rather large in the two sets of coefficients.

5. Final Estimate of the Elasticities and Use of the Results

Comparing the sum $(q_a + q_c + q_f)$ and q_t and q_t in (3) and using a theorem on cross elasticities (the Hotelling-Jurén relation),¹⁸ it is possible to obtain some information about the values of the cross elasticities in (3) not yet estimated. Though no exact solution can be arrived at, it seems to be a safe conclusion that the elasticities $e(a, f)$, $e(c, f)$, and $e(f, c)$ are very small or nil, while $e(f, a)$ has a positive value, probably around 0.2. However, the last estimate is a very crude one, and results obtained later by using this value are therefore put in brackets.

As we have seen before, the cross elasticity $e(a, c)$ and the price elasticity $e(c, c)$ have very small values; in practice they can be neglected, and we then get the following final estimates of the functions giving the demand for the three aggregates of food items:

$$\begin{aligned} q_a &= k_a p_a^{-0.45} \\ q_c &= k_c p_a^{0.40} \\ q_f &= k_f p_a^{0.20} p_f^{-0.80} \end{aligned} \quad (7)$$

Table 10 shows some estimates of the changes in the demand for foods when prices are reduced. The estimates are based on the relations (7).

The estimates refer to the northwestern European region. As an example based on Table 10, a 20 per cent reduction of the prices of fruit and vegetables is expected to be followed by a 20 per cent increase in the consumption of this food category. The average annual per capita consumption in this category amounts to around 45 kg. of fruit and 75 kg. of vegetables; total consumption is about 6.3 million tons of fruit and 10.4 million tons of vegetables.¹⁹ The illustrative price reduction—20 per cent

¹⁸ See Wold-Jurén, *loc. cit.*, p. 112.

¹⁹ Source: FAO food balance sheets. The area is represented by the following countries: United Kingdom, France, Western Germany, the Netherlands, Belgium, Ireland, Denmark, Norway, Finland, and Sweden.

TABLE 10
ESTIMATED CHANGES IN DEMAND AT CERTAIN PRICE REDUCTIONS

Price Reduction in:	Demand for:	Percentage Changes in Demand for Price Reductions of		
		10%	20%	30%
Animal foods.....	{ Animal foods	+5	+11	+17
	{ Cereals and potatoes	-4	-9	-13
	{ Fruits and vegetables	(-2)	(-4)	(-7)
Fruits and vegetables....	{ Fruits and vegetables	+9	+20	+33
Animal foods and fruits and vegetables.....	{ Animal foods	+5	+11	+17
	{ Cereals and potatoes	-4	-9	-13
	{ Fruit and vegetables	(+7)	(+15)	(+24)

—is expected to cause an increase in total demand of about 3 million tons, comprising 1.7 million tons of fruit and 1.3 million tons of vegetables.

The estimate for vegetables is, however, scarcely a realistic one, mainly because the consumption in France is already very high. If no further increase in the demand for vegetables is assumed to be possible in that country, the total consumption increase in the rest of the area amounts to 1.0 million tons in the example given.

If use is made of Table 10 and other tables, numerous other estimates and forecasts can easily be made. It is quite obvious, however, that primary material available for inter-country comparisons is still far from sufficient, and forecasting in quantitative terms is, therefore, rather difficult. On the other hand, results of the kind recorded above can be used as an aid in order to strengthen qualitative judgments of a common type.

6. Conclusions

In short, the results in Section 2 above indicate that in the northwestern European countries the average price elasticity of the demand for animal foods has a significant value, about -0.4 or -0.5 , but that the demand for animal foods is practically independent of changes in cereal prices. On the other hand, the price elasticity for cereals and potatoes is almost zero, but the dependence on animal prices cannot be disregarded, the cross elasticity having a value around 0.4 . Fruit and vegetables are rather sensitive to price variations, the elasticities being around -1.0 and -0.5 . Total caloric intake is almost entirely independent of price changes (elasticity 0.0 to -0.1), but this is not the case when total demand is defined as a volume measured in constant prices, the elasticity then being -0.20 to -0.25 .

It is worth while to compare the price elasticities now referred to (in the case of cereals, the cross elasticity) with the income elasticities in Section 1 (Table 5), at the corresponding average income level (somewhere around \$200). The values for animal products, cereals, total caloric intake and total food volume—obtained by the use of different sets of data—

are all in close agreement, as they should be.²⁰ This offers valuable support for the reliability of the empirical findings.

To sum up, we may conclude that the relationship is rather close between food consumption on the one hand and retail food prices and consumers' income on the other. As far as western Europe is concerned, price and income have a primary influence on the pattern of food demand, not

TABLE 11
BASIC DATA USED FOR TABLE 5 AND FIGURES 1-3*

Country	Food Consumption						"Income"‡ in U.S. Dollars per Head			
	Calorie Intake		Cereals		Animal Products		Pre-war	Postwar		
	Pre-war	Post-war	Pre-war	Post-war	Pre-war	Post-war		Alt. II		
								Alt. I	A	B
Greece.....	2,605	2,500	1,567	1,482	292	238	50	128	40	120
Poland.....	2,710	2,620	1,274	1,389	607	445	63	300	107	330
Hungary.....	2,772		1,565		722					
Italy.....	2,510	2,380	1,571	1,471	343	369	79	235	87	260
Finland.....	2,999	3,115	1,238	1,280	919	1,103	115	348	131	390
Austria.....	2,993	2,670	1,313	1,233	950	749	123	216	127	380
Ireland.....	3,392	3,475	1,316	1,282	1,183	1,268	142	420	168	500
Belgium.....	3,003	2,890	1,294	1,009	731	968	153	582	168	500
France.....	2,884	2,770	1,205	1,167	687	762	153	482	168	500
Netherlands.....	3,007	3,030	978	942	842	818	159	502	175	520
Norway.....	3,158	3,165	1,172	1,117	881	1,266	175	587	219	660
Denmark.....	3,416	3,180	912	959	1,133	1,178	183	689	210	630
Germany†.....	2,961	2,755	1,071	1,075	968	758	195	320	187	560
U.K.....	3,097	3,090	941	986	1,103	1,066	226	773	233	700
Sweden.....	3,122	3,225	932	838	1,086	1,355	227	780	288	860
Switzerland.....	3,106	3,215	1,096	1,119	1,093	1,045	227	849	261	780
U.S.....	3,164	3,175	893	772	1,133	1,270	300	1,453	485	1,450

* In general, prewar consumption data refer to 1934-38, postwar data to 1949/50-1950/51. The prewar figures for income refer to 1938, postwar figures to 1950/51.

† Postwar data refer to Western Germany.

‡ See text.

Sources: Consumption: FAO *Food Balance Sheets*; income: *Economic Survey of Europe in 1949* (ECE, Geneva, 1950), *National and Per Capita Incomes of Seventy Countries in 1949* (UN, October, 1950), and national statistics.

only within countries through time but also between countries, and probably also between classes of the population.

The actual development since 1934-38 in western European food consumption is summarized below and followed by a judgment regarding possible future trends.

1. At gradually rising incomes a substantial improvement in food

²⁰ In theory, for the individual demand of any commodity, the income elasticity (taken with opposite sign) equals the sum of price and cross elasticities (the Slutsky-Schultz relation).

standards is still the normal feature (higher consumption of animal products and other expensive items such as fruit and vegetables²¹ at the expense of cereals and potatoes).

2. Since 1934-38 this long-term tendency has been concealed in most countries mainly because inter-food price relations have shown extraordinary changes.

3. Yet, the long-term tendency is apparent in countries with rapidly growing income per head.

4. With rising income in the future an improvement is expected even if the inter-food price relations remain unchanged.

TABLE 12
CONSUMPTION OF MAIN FOOD ITEMS, GRAMS PER CAPITA PER DAY, IN 1950/51

Country	Pork	Other Meat	Eggs	Milk and Cheese*	Butter	Cereals	Potatoes	Fruits†	Vegetables	Population (millions)
Austria.....	57	47	13	457	8	349	294	135‡	167	7.0
Denmark.....	87	66	23	638	16	268	343	132	192	4.3
France.....	45	107	30	403	16	323	343	95	378	43.1
Western Germany.....	53	48	20	482	17	276	504	128	135	50.0
Italy.....	9	34	18	273	4	420	86	146	241	46.4
Netherlands.....	47	41	12	630	8	278	350	139	190	10.2
Norway.....	38	61	20	927	11	319	347	74	76	3.3
Sweden.....	69	72	28	839	33	251	318	150	69	7.1
Switzerland.....	50	75	24	839	16	325	228	274	217	4.7
United Kingdom.....	33	78	36	606	20	276	302	132	136	50.0

* In milk equivalent.

† Average 1949/50-1951/52.

‡ Average 1949/50-1950/51.

Source: *FAO Food Balance Sheets*.

5. This improvement would be strengthened if price relations returned towards the pre-war ones; such a tendency has, in fact, appeared in recent times.

6. A still stronger improvement would follow if the general food price level were to become lower than at present.

7. An improvement in food standards no longer calls for an appreciable increase in calorie supply (i.e., total calorie supply need not rise much more than is required to keep pace with population growth). But there will still be a considerable increase in the per capita demand for food expressed in constant prices.

APPENDIX

NOTES ON SOURCES AND METHODS

TABLE 1: Real income and actual income elasticities referring to Sweden are taken from *Demand Analysis*, Table 16.6.1. Corresponding income levels in

²¹ Sugar, although a fairly cheap item, belongs to this group.

Figure 1 are obtained under the assumption that real income of industrial workers and low-grade employees was equal to the pre-war average income level in Sweden. Smoothed elasticity values were obtained by determining the Törnqvist function that gives the "best" elasticities as compared with the actual values. The plotted values for population classes in Figure 1 belong to this function. Elasticities obtained from Figure 1 are computed from the first formula in (1).

TABLE 2: *Sources. Twelfth International Dairy Congress*, Section IVb, Subject 3, Stockholm, 1949, with the exception of figures for 1952 which have been computed from Swedish statistics.

TABLE 3: *Sources. (a) The National Food Situation*, July, 1944 for 1909/16-1935/39; (b) *FAO Food Balance Sheets* for 1935/39-1951/52.

TABLE 4: *Sources. Demand Analysis*, Tables 17.5.1, 17.6.2, 17.7.1 and 2. As to the elasticities from multi-country curves, see note to Table 5.

TABLE 13
RETAIL PRICES PER KG OF MAIN FOOD ITEMS, IN OCTOBER, 1951
(Average Price of Cereals = 1)

Country	Pork	Other Meat	Eggs	Milk and Cheese	Butter	Cereals	Potatoes	Fruits	Vegetables
Austria (Vienna).....	4.3	5.6	5.4	0.44	7.4	1*	0.34	1.5	0.46
Denmark (Copenhagen).....	6.7	5.2	6.4	0.53	6.6	1*	0.37	1.8	0.57
France (Paris).....	7.2	6.1	5.9	0.58	10.4	1	0.23	1.9	0.55
Western Germany.....	6.2	5.3	6.2	0.51	8.2	1*	0.21	2.0	0.55
Italy (Rome).....	8.0	7.3	5.8	0.71	11.9	1	0.32	0.9	0.27
Netherlands (Amsterdam).....	9.0	9.5	8.8	0.50	10.7	1	0.38	1.9	0.67
Norway (Oslo).....	9.2	7.3	8.2	0.61	8.9	1*	0.37	3.0	1.15
Sweden (Stockholm).....	5.1	3.7	4.1	0.37	6.0	1*	0.34	1.6	0.68
Switzerland (Bern)....	10.7	7.6	7.8	0.68	13.5	1	0.42	1.5	0.57
United Kingdom (7 cities).....	7.2	6.8	9.0	1.21	8.3	1	0.51	2.2	1.16

* Wheat flour ($\frac{1}{2}$), wheat bread ($\frac{1}{2}$), rye bread ($\frac{1}{2}$)

Source: ILO *Yearbook*, 1951/52.

Note: The following price quotations and weights have been used for aggregated items. Pork: chops ($\frac{1}{2}$), shoulders ($\frac{1}{2}$); Other meat (beef): sirloin ($\frac{1}{2}$), brisket ($\frac{1}{2}$); Milk and cheese (in milk equivalent): milk price only: Cereals: wheat flour ($\frac{1}{2}$), wheat bread ($\frac{1}{2}$); Fruits: Apples ($\frac{1}{2}$), oranges ($\frac{1}{2}$); Vegetables: Cabbage ($\frac{1}{2}$), onions ($\frac{1}{2}$).

TABLE 5: Pre-war elasticities for animal foods, all other food items, and total food in calories are computed by use of the formulas (1). The elasticities for cereals are obtained in the same manner as those for "all other food items," i.e., by taking the difference between two functions, in this case the curve fitted to data for total food and a curve for total food except cereals. Elasticities, $E_c(r)$, for total food volume, $q_c(r)$, are computed by using the weights of 5 for animal foods and 1 for all other food items.²² The formulas used were:

²² As an average (which varies from country to country) the retail price value of one calorie of animal foods is about five times that of one calorie of other food items. See also *European Agriculture* (Geneva, 1954), Chart 4, showing relation prices between 3 and 4. These lower values are explained by the fact that the relatively cheap products sugar, vegetable oils, and fats are excluded in the chart.

$$E_c(r) = \frac{E_t(r) \times q_t(r) + 4E_a(r) \times q_a(r)}{q_t(r) + 4q_a(r)}$$

$$q_c(r) = 5q_a(r) + q_o(r) = q_t(r) + 4q_a(r)$$

The price ratio used is, of course, a very crude measure. However, even a rather large variation in the price ratio does not affect the elasticities very much as can be seen by the following table:

Income Level	Elasticities Obtained by Using Weights		
	4:1	5:1	6:1
50.....	0.39	0.42	0.46
125.....	0.28	0.31	0.34
300.....	0.18	0.20	0.21

The post-war elasticities are computed in a similar manner. (The basic data used for Table 5 and Figures 1-3 are shown in Table 11.) However, two sets of post-war income data have been applied:

Alternative I: National income per capita in current U.S. dollars.

Alternative II: The pre-war data used (total commodities available per head in 1938 U.S. dollars) are multiplied country by country by indices of real national income per head (Alternative IIA). In Alternative IIB these figures are multiplied by the factor 3.0 in order to get the data on a level with Alternative I. Post-war income in Figures 2 and 3 refer to Alternative IIA.

TABLE 6: *Sources.* (a) FAO *Food Balance Sheets* for calorie intake.
(b) Tables 12 and 13 for retail price values.

TABLE 9: Consumption at constant income level has been computed by aid of relations (1). As to fruit and vegetables, the income elasticities quoted in the previous text have also been used.

Psychological and Objective Factors in the Prediction of Brand Choice; Ford versus Chevrolet*

.....

FRANKLIN B. EVANS†

I. INTRODUCTION‡

IN RECENT YEARS A NUMBER OF NONQUANTITATIVE STUDIES IN MARKETING have found substantial differences in the personalities of owners of different automobile makes. Buyers of one brand are described as differing sharply, personality-wise, from those of another. Also, the brands themselves are thought to have images or personalities extending beyond their physical characteristics. These images are expected to draw buyers, often in terms of personality need satisfaction.

This study was undertaken to test the ability of psychological and objective methods to discriminate between owners of the two largest-selling automobiles, Ford and Chevrolet. The cars are objectively almost perfect substitutes; their prices, models, and other features are almost identical. However, previous research has indicated that these makes represent different psychological images to the public and that the purchasers of one make are sharply different, psychologically speaking, from purchasers of the other, at least on the average.

A simple random sample of Ford and Chevrolet owners provided the basic data for the test. The owners' scores on a standard test of manifest psychological needs were used as a basis for judging the ability of psy-

* Reprinted from the *Journal of Business*, Vol. XXXII, No. 4 (October, 1959), pp. 340-69 by permission of the University of Chicago Press. Copyright 1959 by the University of Chicago, all rights reserved.

† University of Chicago.

‡ The study resulting in this publication was in part made under a fellowship granted by the Ford Foundation. This financial aid is gratefully acknowledged. However, the conclusions, opinions, and other statements in this article are those of the author and are not necessarily those of the Ford Foundation. The writer also wishes to express his appreciation to Drs. Harry V. Roberts, Morris I. Stein, and James S. Coleman for their counsel and criticisms throughout the preparation of this paper.

chological factors to predict the brand of car owned. Demographic and other objective factors were also obtained, and thus predictive power was measured. These variables represent two widely different approaches to market research. Manifest psychological needs may be said to represent the motivations research approach, while the objective factors typify a more traditional approach which emphasizes the economic and demographic variables influencing the demand curve.

In each class of variables some small and only barely statistically significant differences were found between Ford and Chevrolet owners. These differences, however, are too minor to use effectively in predicting the brand of car owned. Taken singly or in a linear combination, neither personality needs nor demographic variables assigned brand ownership with any considerable degree of certainty. Even the advantage of selecting the most predictive variables from each class and combining them into a single linear discriminant function did little to improve the predictive efficacy.

The bulk of the literature ascribing differences to Ford and Chevrolet owners comes from the motivation researchers. When a linear relationship of the personality variables failed to produce the desired discrimination, psychologists suggested that perhaps some other model would fit better. However, no nonlinear relationship of the personality needs could be discovered. In addition, a select group of psychologists was unable to assign the brand correctly on the basis of the need scores. Also, neither grouping the needs according to type of basic satisfaction involved nor examination of the ranges of their scores showed any important difference between Ford and Chevrolet owners.

Two subsidiary analyses of other aspects of brand choice proved no more fruitful. Brand images were found, but they were much more diffuse than others have indicated. In only five out of twenty-one cases did they show images in a rigorous sense. The loyal and non-loyal owners of each brand are much alike with respect to the most predictive of the independent variables.

Market Research Methods

Traditional Research. Traditional market researchers have stressed the importance of objective variables. Implicit in the use of these objective and demographic variables is their importance as demand determinants. These researchers have also relied upon consumers' opinions and motives that can be verbalized in response to direct questions. They believe that consumers both can and will disclose the reasons for, and thus predict, their behavior. Sample sizes tend to be large, sometimes running to thousands. Traditional researchers have used, or at least advocated, standard statistical techniques.

Variables such as age, income, race, sex, or geographic location are used to describe consumers. The narrower the ranges of these variables, the

better a particular market can be distinguished from other competing products. Besides describing the limits of a market, these variables can often be used to explain and predict purchase behavior.¹

Motivation Research. Whereas most traditional researchers have had business, economic, or statistical training, motivation researchers have entered the field with behavioral science backgrounds. Most have training in psychology or social psychology, but anthropology, psychiatry, psychometrics, and sociology are also represented.

Much of the work of motivation researchers has contained the tacit assumption that common motivations exist for large segments of the population. Standard statistical techniques are seldom employed; samples are usually small, rarely exceeding a few hundred or even a few score. Survey respondents are picked either to fit a predetermined quota or to be representative of a particular social class, and the actual selection of the individual to be interviewed is often left completely to the field worker's discretion.

Drawing heavily upon the Freudian schools of psychology, motivation researchers have maintained that many purchase reasons are deeply rooted in personality. These motives are either unknown to the conscious self or too ego-threatening to be revealed by direct questioning. To uncover these motives, depth interviews are used. In addition to depth interviewing, most motivation researchers use some kinds of psychological testing.

Although the value of motivation research has been questioned for some years now, the dispute has seldom been more than polemic. Few actual experiments have been made to compare the recommendations of the different schools.² An indirect measure of the predictive value of motivation research was tested by the writer. It was found that the recommendations of at least one study were predictive of advertising readership at a significant level.³ There is an over-all lack of evidence comparing the two research modes.

Psychological Testing

Almost all motivation researchers use some kinds of psychological tests, ranging from sentence completions and adjective checklists to completely unstructured projectives like the Rorschach ink blots. As there is very little standardization in this area or agreement about the instrument

¹ Mordechai E. Kreinen, John B. Lansing, and James N. Morgan, "Analyses of Life Insurance Premiums," *Review of Economics and Statistics*, Vol. XXXIX (1957), pp. 46-54.

² See, however, John Masek, "A Study of the Usefulness of Motivation Research in Improving the Efficiency of Direct-Mail Selling Efforts" (unpublished Ph.D. dissertation, School of Business, University of Chicago, 1957).

³ Franklin B. Evans, "Motivation Research and Advertising Readership," *Journal of Business*, Vol. XXX (1957), pp. 141-46.

used, a group of experts were questioned.⁴ These experts are either well-known practitioners in the field or academic people who have studied and written on the subject.

Eleven out of the sixteen experts believe psychological tests to be valuable and necessary adjuncts to motivation research. Of the five who did not approve their use, only one is actively engaged in the field. The five dissenters questioned the over-all value of psychological tests, not only their marketing uses.

Of the eleven experts favoring psychological tests, eight stated that the tests could be used in their clinical form. Others suggested that the tests should be modified to relate directly to product attitudes. Thus a test's basic form and presentation are maintained, but the context is modified to relate views focused upon a particular product or brand.

Seven of the experts believed that motivation research techniques are standardized enough for others to replicate them. The question is of importance because, unless replication is possible the results of a particular study hinge entirely upon the skill and intuitiveness of the analyst. Many of the critics of motivation research have stressed this point. Some also mentioned that even the psychological tests like the Rorschach and Szondi are not very well standardized in clinical use.⁵

Research on Automobiles

Discrimination of Brand Purchasers. Referring to automobiles, Pierre Martineau, director of research and marketing of the *Chicago Tribune*, recently wrote: "The buyers and non-buyers were undistinguishable except on a personality basis."⁶ The same general view is often publicly expressed by executives of these companies. Henry G. Baker of the Ford Motor Company has said: "A make thus becomes a very real extension of the owner's DESIRED personality."⁷ And, referring to the symbolic as-

⁴ Those replying were: Wroe Alderson (Alderson and Sessions, Inc.), Seymour Banks (Leo Burnett, Inc.), George H. Brown (Ford Motor Company), Louis Cheskin (Color Research Institute), Ernest Dichter (Institute for Motivational Research), Robert Ferber (University of Illinois), Burleigh B. Gardner (Social Research, Inc.), Mason Haire (University of California), William E. Henry (University of Chicago), Herta Herzog (McCann-Erickson, Inc.), Darrell B. Lucas (New York University), Joseph W. Newman (Harvard University), Fred T. Schreier (A. J. Wood and Company), George Horsley Smith (Rutgers University), W. Lloyd Warner (University of Chicago), and Hans Zeisel (University of Chicago). Also, although they did not answer the specific questions asked (and hence are not included in the survey results), replies of a general nature were received from Stuart H. Britt (Northwestern University), Paul F. Lazarsfeld (Columbia University), and James M. Vicary (James M. Vicary Corp.).

⁵ Mason Haire, personal correspondence, February 14, 1958.

⁶ Pierre Martineau, *Motivation in Advertising* (New York: McGraw-Hill Book Co., Inc., 1957), p. 67. Martineau's observations were based upon a study made in 1954 by Social Research, Inc., for the *Chicago Tribune*.

⁷ Henry G. Baker, "Sales and Marketing Planning of the Edsel," in Robert L. Clewett (ed.), *Marketing's Role in Scientific Management* (Chicago: American Marketing Association, 1957), p. 130.

pects of automobiles, David Wallace of the same company said: "On this dimension Ford is perceived as being the most masculine of the low-priced makes. Chevrolet and Plymouth are more feminine."⁸

Although Ford and Chevrolet purchasers have comprised about half the automobile market in the last twenty-five years, they are commonly portrayed as being entirely different personality-wise. A composite of these descriptions pictures Ford owners to be independent, impulsive, masculine, alert to change, and self-confident, while Chevrolet owners are described as conservative, thrifty, prestige-conscious, less masculine, and seeking to avoid extremes.

Criticisms. During the recession of 1958 the sales of most domestic automobiles fell sharply, while the sales of small imported cars increased. The American manufacturers were blamed for producing cars which the public did not want. One of the most vituperative criticisms came from semanticist S. I. Hayakawa. He stated: "The trouble with car manufacturers (who, like other isolated people in undeveloped areas, are devout believers in voodoo) is that they have been listening too long to the motivation research people."⁹ A similar, if less caustic, view of research on automobiles was expressed by Harold E. Churchill, president of the Studebaker-Packard Corporation. Referring to Studebaker's forthcoming small car, he said: "It is not a car based on a small sample survey of the social significance of the automobile today."¹⁰

Study Design and Implementation

Research Strategy. With limited financial resources it was not possible to study all the various kinds of people who own Fords and Chevrolets. Therefore, a restricted and relatively homogeneous group—residents of Park Forest, Illinois—was selected for study.¹¹ The purpose of this limited study is to (1) demonstrate an improved methodology and, more important, (2) give limited but well-founded results that are a challenge to others.

The universe was further restricted to Ford and Chevrolet owners of 1955–58 models. This was done to minimize the effects of style cycles of these brands and includes model years in which each was the top seller nationally. In addition, all owners are white males and have only one car. From this restricted universe a simple random sample was drawn.

By confining the sample to this limited universe, it is believed that more sensitive discrimination will be possible, especially in terms of the

⁸ David Wallace, "An Adventure in People's Minds: Finding a Personality for the E-Car," in Stewart H. Rewoldt (ed.), *Conference on Sales Management* ("Michigan Business Papers," No. 14 [Ann Arbor: University of Michigan, 1957]), p. 6.

⁹ S. I. Hayakawa, "Irrational Dreams, but Rational Behavior," *Advertising Age*, May 12, 1958, p. 111; reprinted from *ETC: A Review of General Semantics*, Spring, 1958.

¹⁰ *Road and Track*, July, 1958, p. 2.

¹¹ See Appendix, "The Park Forest Universe," for the rationale for this choice.

personality variables analyzed. Gross differences such as sex or social class could confound the results in a sample of limited size.

The findings of this study cannot be generalized statistically to populations other than Park Forest, Illinois. However, it is believed that this test of the discriminatory efficacy of psychological and demographic variables will provide knowledge germane to the larger problems of prediction of brand choice which have seldom been solved. To say it differently, the inferences that can and should be made to populations broader than Park Forest must be based on marketing judgment rather than on conventional statistical inference.

Test Instrument. The questionnaire was designed to collect three specific kinds of data—demographic and factual data related to automobile ownership, role-playing questions designed to measure perceived differences of Ford and Chevrolet owners, and psychological needs reflecting the respondents' basic personalities. Data for the first two categories were taken in personal interview. Personality needs were measured by a paper-and-pencil test filled out by the respondent in the interviewer's presence.

The collection of data in Park Forest took place in June and July, 1958. One hundred and forty-six substantially completed interviews were secured; 140 on the psychological test.¹²

Analysis of the Data. The major purpose of the analysis is to discover which variables best predict brand ownership. Several methods are presented, but primary reliance is placed upon the linear discriminant function.¹³ Statistically, this function reduces the multivariate problem to a univariate one. A linear equation is derived which maximizes the separation between the two groups by optimum weighting of the independent variables. This equation is of the following form:

$$Y = aX_1 + bX_2 \dots kX_k$$

The X_1 , X_2 , etc., represent the independent variables, and a , b , etc., their weights. Conventionally the first weight (a) is made unity, and the others are expressed in proportionate terms. Substitution of the group means for each variable in the equation and solving it (summation) yields a numerical index ($\bar{Y}_i$), where $i = 1$ or 2 , which describes the group. Substitution of an individual's scores gives an index (Y) for him. Any new individual may then be assigned to the group ($\bar{Y}_1$ or $\bar{Y}_2$) whose score is closest to his.

II. PERSONALITY FACTORS

The Psychological Test

Instrument. The analysis of personality variables presented in this study is based upon the results of a psychological schedule filled out by

¹² See Appendix, "Data Collection," for further information on response factors.

¹³ R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems," *Annals of Eugenics*, Vol. VII (1936-37), pp. 179-88.

seventy-one Ford owners and sixty-nine Chevrolet owners. The test was constructed from items in the Edwards Personal Preference Schedule.¹⁴ This test purports to measure manifest personality needs as described by Murray.¹⁵ This is a simple paper-and-pencil test consisting of sets of paired statements in which each sentence in a pair describes a personality need. From each pair the respondent selects the statement he feels best portrays himself. One hundred and ten sets of paired comparisons were used, somewhat fewer than the full test because of interviewing problems encountered in pretesting. These paired comparisons yielded scores for each of eleven personality needs.

As previously indicated, most of the emphasis in motivation research has been upon projective tests—tests in which the respondent is presented with an ambiguous picture or situation and an answer in this situation reveals something about the personality. A simple paper-and-pencil test was used in this study in preference to a projective test for several practical reasons.

Projective tests are time-consuming and costly to administer. They require interviewers specifically trained in their use. To secure the response necessary for this study would have increased the project's cost several fold or diminished the sample size to the point where only the grossest differences could be demonstrated. Besides cost factors, the scoring and interpretation of projective tests are both difficult and unstandardized. Their use would have entailed securing several judges to rate (and agree upon) each respondent's personality pattern.

The Edwards Personal Preference Schedule was chosen for the following reasons: (1) scoring is simple, mechanical, and unambiguous; (2) it is gaining wide use among psychologists, and published results are available for comparison purposes; (3) it is based upon Murray's system of personality needs. The same needs are used in the Thematic Apperception Test (TAT), the most popular of the projective tests. For these reasons the Edwards was chosen for this analysis, although two other personality tests were briefly tried in the pretest stage of the project.

Personality Needs Measured. Psychologists believe that to a considerable extent a person is known by his needs.¹⁶ The pattern of needs in a personality defines the individual and allows something meaningful about him to be communicated to others. To say that a person is exceptionally aggressive, for example, characterizes him in a very general way. The needs treated as psychological variables in this paper are as follows:¹⁷

¹⁴ Allen L. Edwards, *Edwards Personal Preference Schedule Manual* (New York: Psychological Corp., 1957).

¹⁵ Henry A. Murray *et al.*, *Explorations in Personality* (New York: Oxford University Press, 1938), pp. 142-242.

¹⁶ Harold J. Leavitt, *Managerial Psychology* (Chicago: University of Chicago Press, 1958), p. 98.

¹⁷ Edwards, *op. cit.*, p. 14.

1. *Achievement*: To do one's best, to accomplish something of great significance.
2. *Deference*: To find out what others think, to accept the leadership of others.
3. *Exhibition*: To say witty and clever things, to talk about personal achievements.
4. *Autonomy*: To be able to come and go as desired, to say what one thinks about things.
5. *Affiliation*: To be loyal to friends, to make as many friends as possible.
6. *Intracception*: To analyze one's motives and feelings, to analyze the behavior of others.
7. *Dominance*: To be a leader in the groups to which one belongs, to tell others how to do their jobs.
8. *Abasement*: To feel guilty when one does something wrong, to feel inferior to others in most respects.
9. *Change*: To do new and different things, to participate in new fads and fashions.
10. *Aggression*: To attack contrary points of view, to get revenge for insults.
11. *Heterosexuality*: To become sexually excited, to be in love with someone of the opposite sex.

Independence of Needs. In order to use these need scores as independent measures of personality, examination was made of their intercorrelations. High intercorrelations among them would indicate that the items do not measure independent personality dimensions. For his normative group of 1,509 college men and women, Edwards found relatively low intercorrelations.¹⁸ For the sample of 140 Ford and Chevrolet owners, intercorrelations of the need scores ranged from +.255 to -.364.¹⁹

Test Scores of Ford and Chevrolet Owners

Group Means. The average score for each of the personality needs for Ford and Chevrolet owners is shown in Table 1. For seven of the needs (achievement, deference, intracceptions, abasement, change, aggression, and heterosexuality) the scores show no statistically significant difference. Three other needs (exhibition, autonomy, and affiliation) are significantly different at about the 10 per cent level. Only for dominance do the groups differ beyond the 5 per cent level of significance. A two-tailed test was used to test for differences between means because, from the orientation of this study, there were no specific hypotheses indicating the direction of the differences. However, all the differences except achievement and autonomy are in the direction commonly indicated by motivation researchers. Ford and Chevrolet owners as groups agree closely in the rank order they place the needs.²⁰ To put it differently, the differences between mean scores for different needs are much larger than

¹⁸ *Ibid.*, p. 17.

¹⁹ For further discussion of the test's reliability see Appendix, "The Psychological Test."

²⁰ Rank-order correlation = 0.903. Reject the null hypothesis at 0.01 level of significance.

TABLE 1
AVERAGE PERSONALITY NEED SCORES OF FORD
AND CHEVROLET OWNERS

	<i>Ford</i> (<i>N</i> = 71)	<i>Chevrolet</i> (<i>N</i> = 69)	<i>Difference</i>
Achievement.....	12.80	12.87	-0.07
Deference.....	9.47	9.77	-0.30
Exhibition.....	10.06	9.30	+0.76*
Autonomy.....	7.86	8.80	-0.94*
Affiliation.....	10.14	11.09	-0.95*
Intracception.....	11.17	11.32	-0.15
Dominance.....	13.69	12.41	+1.28†
Abasement.....	7.20	7.28	-0.08
Change.....	11.39	11.06	+0.33
Aggression.....	9.59	9.52	+0.07
Heterosexuality.....	6.63	6.59	+0.04

* Significant at 10 percent level (two-tailed test).

† Significant at 5 percent level (two-tailed test).

the differences between Ford and Chevrolet means for the same need. Moreover, the differences between Ford and Chevrolet means, even on dominance, which showed the greatest separation, are of slight value for predicting a person's brand selection. The distributions of scores for all needs overlap to such an extent that discrimination is virtually impossible.

Graphic Analysis. Affiliation and dominance are the two needs which showed the greatest differences in group means. Figure 1 shows a cross-classification of these needs. The overlap of the distributions and their lack of discriminatory ability is apparent. Examination shows little clustering by brand, and there are no visible patterns that would indicate any simple way for predicting the brand owned. Similarly, Figure 2 shows abasement and aggression—needs whose group means were almost identical. As in Figure 1, the lack of discrimination by these needs is obvious. Although fifty-five such graphic cross-classifications could be made of the eleven needs, it is believed that these two (Figs. 1 and 2) demonstrate the problem. They support the earlier analysis of individual need scores; considering needs in pairs seems to add little or nothing to predictive ability.

Linear Discriminant Function

Statistical Analysis. To test for discrimination between Ford and Chevrolet owners using ten need scores (ten independent variables) at one time, a linear discriminant function was computed. The purpose is to weight the need scores of the two contrasted groups to provide maximum linear separation between them. That is, we restrict ourselves to a model that computes predicted scores by linear equations and finds the "best" coefficients for these equations in achieving discrimination.²¹

²¹ Fisher, *op. cit.*, pp. 179-88. See Appendix, "Linear Discriminant Function," for more detailed treatment and further references.

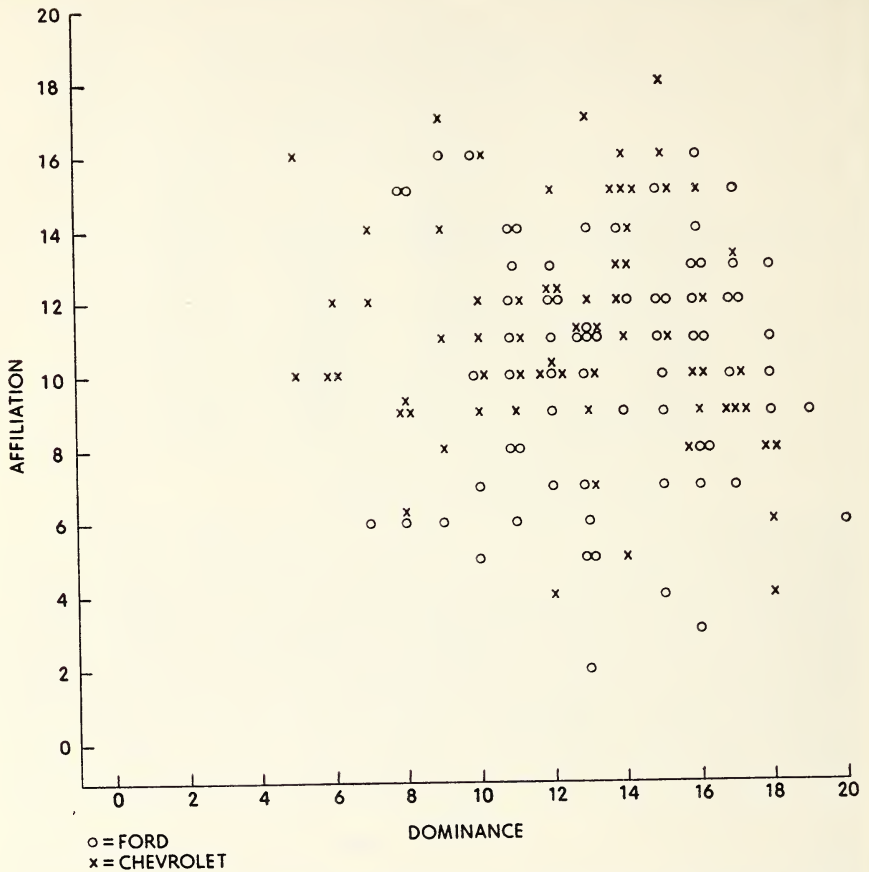


FIGURE 1. Personality needs. Cross-classification of the needs for affiliation and dominance for Ford and Chevrolet owners.

Only ten needs are used in this analysis, to avoid the statistical restrictions of a singular matrix. Sex, the need with the lowest score, was not used explicitly. The score of any individual need could vary from 0 to 20. For the ten needs used, a respondent's score could range from 90 to 110; for all eleven needs his score is necessarily 110.

Weights of the Variables. The weights of the ten psychological variables for the discriminant function are shown in Table 2. Group means and the mean of the Y for each group are also shown.

To test whether this function really discriminates between Ford and Chevrolet owners, an analysis of variance was performed. This is shown in Table 3. The multiple correlation coefficient (R) is .3353, and its square (R^2) is .1124. The resulting F ratio with 10 and 129 degrees of freedom is 1.634. This is just barely significant at the 10 per cent level, indicating that the linear discriminant function of personality needs is of doubtful statistical significance in this case.

TABLE 2
LINEAR DISCRIMINANT FUNCTION OF PERSONALITY NEEDS
FOR FORD AND CHEVROLET OWNERS

Variable	Weight	Group Means	
		Ford (N = 71)	Chevrolet (N = 69)
Achievement.....	+1.0000 X_1	12.80	12.87
Deference.....	-0.0481 X_2	9.47	9.77
Exhibition.....	-1.4058 X_3	10.06	9.30
Autonomy.....	+2.0505 X_4	7.86	8.80
Affiliation.....	+2.0944 X_5	10.14	11.09
Intracception.....	-0.2356 X_6	11.17	11.32
Dominance.....	-2.1090 X_7	13.69	12.41
Abasement.....	-0.5269 X_8	7.20	7.28
Change.....	-1.6005 X_9	11.39	11.06
Aggression.....	+0.2561 X_{10}	9.59	9.52
$\bar{Y}_i = \Sigma \text{ weights times group means} \dots$		-15.5150	-7.3416

Applying this discriminant function to the data from which it was developed, it misclassifies fifty-two individuals, or 37.1 per cent of the sample. A completely random basis of classification, such as flipping a coin, would have misclassified approximately 50 per cent of the sample. Table 4 shows the classifications by brand. Also one would not expect the equation to predict even this accurately for new observations.²²

TABLE 3
ANALYSIS OF VARIANCE OF LINEAR DISCRIMINANT FUNCTION
OF PERSONALITY NEED VARIABLES

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Discriminant function.....	10	0.1124(R^2)	0.01124
Remainder.....	129	0.8876(1 - R^2)	0.00688
Total.....	139	1.0000	
	$F = 1.634$		

Similarity of Correctly and Incorrectly Classified Cases. Comparison was made of those classified correctly and those owners misclassified to see whether other factors could be responsible for the lack of predictive efficacy of the function. Three characteristics other than personality

²² The probability of misclassification for this discriminant function was computed by dividing half the difference between $\bar{Y}_1$ and $\bar{Y}_2$ by the within-sample standard deviation of individual Y 's and finding the probability that a standard normal variable would exceed that number (see Fisher, *op. cit.*, pp. 182-83). For this function, the probability of misclassification is 0.366. For future samples this would be an understatement, since this procedure is a large sample approximation that assumes the sample quantities equal corresponding parameters. The true proportion of error would be higher, but exactly how much is not known.

TABLE 4
CLASSIFICATION OF FORD AND CHEVROLET OWNERS BY
LINEAR DISCRIMINANT FUNCTION OF PERSONALITY
AND NEED VARIABLES

	<i>Classified Correctly</i>	<i>Misclassified</i>	<i>Total</i>
Ford.....	41	30	71
Chevrolet.....	47	22	69
	88 (62.9%)	52 (37.1%)	140 (100%)

needs were selected for this analysis. Age of owner is a demographic factor, intention to purchase the same brand again indicates brand satisfaction, and smokers versus non-smokers may reflect some deeper personality differences than those expressed by the need scores. Table 5 shows these comparisons both for the combined group and for Ford and Chevrolet owners separately.

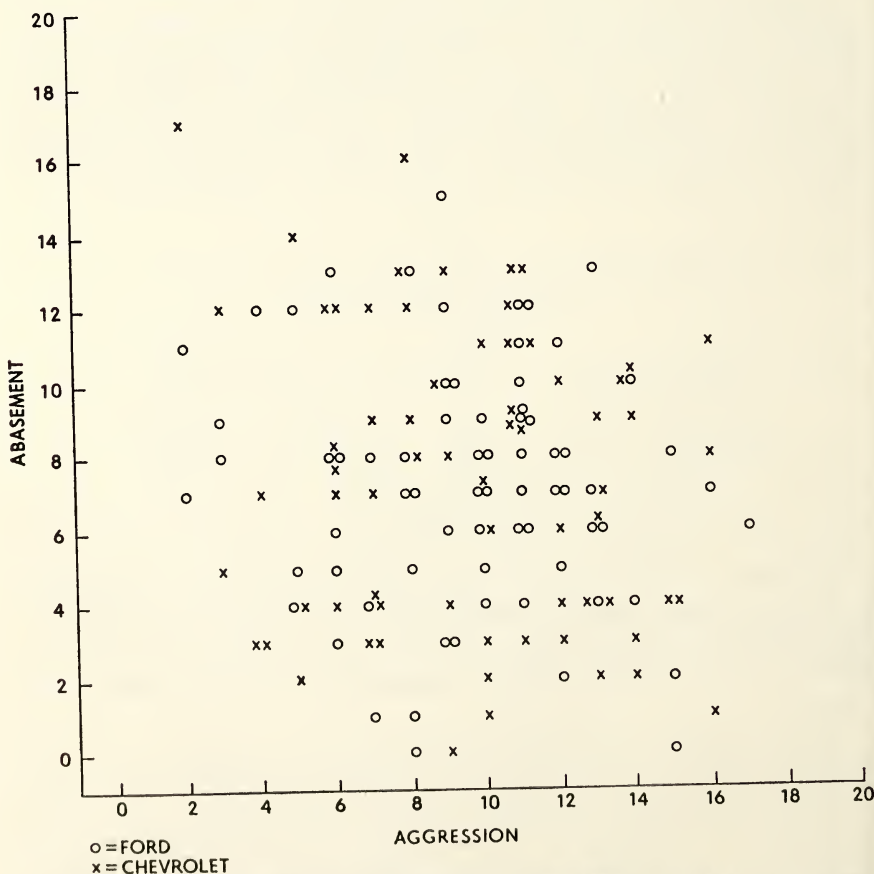


FIGURE 2. Personality needs. Cross-classification of the needs for abasement and aggression for Ford and Chevrolet owners.

TABLE 5

COMPARISON OF FORD AND CHEVROLET OWNERS CLASSIFIED CORRECTLY AND MISCLASSIFIED BY LINEAR DISCRIMINANT FUNCTION OF PERSONALITY NEEDS

Characteristic	Combined Sample		Ford		Chevrolet	
	Classified Correctly (N = 88)	Misclassified (N = 52)	Classified Correctly (N = 41)	Misclassified (N = 30)	Classified Correctly (N = 47)	Misclassified (N = 22)
Age of owners:						
39 years and over.....	20	12	9(22.0)*	8(26.7)*	11(23.4)*	4(18.2)*
31-38 years.....	46	28	22(53.6)	15(50.0)	24(51.0)	13(59.1)
Under 31 years....	22	12	10(24.4)	7(23.3)	12(25.6)	5(22.7)
Buy same brand						
again.....	48	29	18(43.9)	17(56.7)	30(63.8)	12(54.6)
Not buy same brand again.....	40	23	23(56.1)	13(43.3)	17(36.2)	10(45.4)
Owner smokes.....	61	36	33(80.5)	22(73.3)	28(59.6)	14(63.6)
Does not smoke.....	27	16	8(19.5)	8(26.7)	19(40.4)	8(36.4)

* Within-subgroup percentages.

Examination of these characteristics shows that those classified correctly and those misclassified are very similar, both for the combined sample and for each make separately. Tests of the largest differences for each characteristic show no significant difference for either make, even at the 10 per cent level. Also the direction of the differences within brands is often inconsistent when both brands are examined. For example, misclassified Ford owners are more satisfied with their cars than are correctly classified Ford owners; the reverse is true for Chevrolet owners.

This analysis shows that there is no readily apparent explanation for the lack of "fit" of the linear discriminant function. The incorrect classification of almost two-fifths of the sample from which the function was derived shows that the variables are of low predictive value in this case and that a linear combination of ten personality needs is not sufficient to achieve much discrimination between owners of Fords and Chevrolets. Certainly, gross differences are simply not to be found.

Non-Linear Relationships

Psychologists, when assessing personality, attempt to look at the total or whole person as well as the individual components. From this vantage point one would expect psychologists to judge inadequate an analysis of personality in which a simple linear equation is used to express the relationships. The psychologist's assessment of personality is often a subjective process of weighting and weighing all the factors or reactions he has available. Thus, through their own intuitive processes, they may be able to discriminate between personalities in a way that reflects non-linear configurations.

Although the ten needs measured in this study cannot be considered as

"the whole man" if the needs are positively or negatively cathected to automobile ownership, one would expect psychologists to be able to distinguish between two differing groups.

The Psychologists Discriminate. To see whether psychologists using their own particular methods and judgments could distinguish between Ford and Chevrolet owners, a selected group of psychologists was asked to examine the need scores of ten individuals.²³ These eighteen psychologists were given the need scores of five Ford and five Chevrolet owners randomly selected from the Park Forest respondents. The judges were told that there were five Ford and five Chevrolet owners in this group.

Along with the test scores of the ten individuals, these judges were given the following protocol:

In recent years many marketing research studies have either attributed definite personality characteristics to specific kinds of automobiles or categorized brand owners by personality. These are derived from depth-interviewing (unstructured) techniques and psychological testing. Although different methods are employed, results are often surprisingly similar. For example:

Ford owners are said to be more independent, alert to change and experiment, more tolerant, self-confident, impulsive, interested in people and more masculine. They *drive* a Ford, and are younger people with above average incomes.

On the other hand:

Chevrolet owners are more feminine, more cautious, suspicious, conservative, thrifty, prestige conscious, less independent, and hold their cars longer. They *own* a Chevrolet and consider it a stable and dependable car. They want to be up to date but to avoid being too extreme. They are interested in things more than people.

As a group, these judges picked only 70 cases correctly out of 180 possible choices, for a percentage of 39.9. The maximum number of correct choices made by anyone was 6 and the minimum was 2.²⁴

Although they did not correctly assign the owners to their proper groups, these judges did exhibit a high consensus. Sixteen, seventeen, and eighteen judges, respectively, agreed that, of the five Chevrolet cases, three were Fords. Fourteen, fourteen, and fifteen of the judges, respectively, assigned three of the five Ford cases to the Chevrolet group. The three Chevrolet cases judged incorrectly consisted of individuals with high scores for achievement, aggression, and dominance and low scores for abasement. The three Ford cases most misclassified had low scores for achievement, autonomy, and exhibition and a high score for abasement. These six cases accounted for the bulk of the incorrect placements.

When apprised of their results, the psychologists themselves suggested

²³ Eighteen psychologists participated; four are in the academic profession, nine in advertising or marketing research, three in psychological testing and guidance, and two are practicing analysts.

²⁴ The actual distribution of choices and the statistical probability of these occurring owing to chance alone are shown in the Appendix, "Psychologists' Judging by Personality Needs."

several possible reasons for this outcome. The most common explanation offered was that the description of Ford and Chevrolet owners taken from motivation research studies was wrong or misleading. Two pointed out that these descriptions as given are internally inconsistent. The masculine attributes assigned to Ford owners are more in keeping with people interested in *things*, not in *other people* as stated, and the opposite for Chevrolet owners. However, five of the eighteen psychologists acting as judges claimed that they did not follow these protocols; of these five, one made six correct choices and the others four.

The second reason offered was that the test instrument used did not really measure the needs as described. Several of the psychologists said when taking the test that they did not expect to score very well as the data were limited, and one suggested that the psychologists themselves would probably take the test in a very defensive way, thereby distorting results. Still another reason could be that the ten randomly chosen car owners happened to be atypical of their respective groups.

The high consensus of the psychologists on the six cases previously mentioned shows that they were discriminating but apparently not by any subtle process. If a curvilinear relationship of the ten personality needs could do better than the linear discriminant function in classifying the Ford and Chevrolet owners, the psychologists did not discover it either consciously or unconsciously.

Some Further Investigations. The ten needs from the test can be grouped into three general classes.²⁵ Affiliation, abasement, aggression, autonomy, deference, and dominance express interpersonal relations. Exhibition and achievements are inner-state needs, and intraception and change are goal-oriented needs. When aggregated into these three classes, the needs show even smaller differences than when treated separately, as several of the individual differences cancel each other. This again points to the similarity of personality factors of the two brand owners.

Differences in dispersion of the individual need scores between the two groups would also indicate non-linear relationships. A wide range of scores for owners of one brand as opposed to a narrow range for the other would denote different personality types, even though the over-all group means are the same. However, examination of the ranges of the ten needs does not suggest that this exists. For five of the needs the range is the same, and for four more the differences in range are one, one, two, and three, respectively. Only one need—achievement—shows considerable difference in range between the two brands.

Conclusion

All the evidence points to the conclusion that personality needs, as measured in this study, are of little value in predicting whether an in-

²⁵ G. G. Stern, M. I. Stein, and B. S. Bloom, *Methods in Personality Assessment* (Glencoe, Ill.: Free Press of Glencoe, Inc., 1956), pp. 69-73.

dividual owns a Ford or Chevrolet automobile. Although people within a common social class have different personalities, their personalities do not appear to be systematically related to selection of the two most popular brands of cars. This result is, however, based upon the specific test instrument used, and no doubt criticism can be raised on this point. In its defense it can be said that this test is commonly accepted by many psychologists for many different uses.²⁶ Other reasons for its applicability in this particular study have been given previously.

The more important question is beyond the scope of this paper. This is, "Can any test really measure personality?" Psychologists themselves are deeply concerned with this problem and do not hesitate to question the entire area of personality theory.²⁷

III. DEMOGRAPHIC AND OBJECTIVE FACTORS

The Variables

The selection of Park Forest as the area of study restricted the ranges of the demographic variables. In this study the psychological variables showed wide variation, as wide as for other groups to whom this test has been administered. However, in terms of the demographic variables, Park Forest is much more limited. The ages and incomes of the survey respondents, for example, were of much narrower range than would be found in sampling larger and less homogeneous areas. Therefore, the linear discriminant function of demographic variables is under a handicap when compared to the one based upon the psychological needs.

The Variables Selected. Twelve objective variables were chosen from the interview data to represent factors commonly used by traditional market researchers. With the exception of income, all were easily collected. Nine and sixth-tenths percent of the total sample refused to divulge their yearly family income. The twelve variables selected were as follows: (1) age of automobile presently owned; (2) use of automobile more or less than 10,000 miles per year; (3) buyer "shopped" more than one dealer before purchase; (4) smokers versus non-smokers; (5) homeowners versus renters; (6) three or more children living at home; (7) religious preference; (8) church attendance more or less than once a month; (9) political party preference; (10) age of owner; (11) owner has worked for present firm more or less than five years; and (12) family yearly income.

These twelve variables describe several different aspects of the respondents' lives. Model year and usage of the car may reflect the car's importance to the family. Smoking, shopping for new cars, and tenure

²⁶ Robert R. Blake and Jane S. Mouton, "Personality," in Paul R. Farnsworth (ed.), *Annual Review of Psychology* (Palo Alto, Calif.: Annual Review, Inc., 1959), p. 207.

²⁷ *Ibid.*, p. 226.

with a firm may reflect personality measures possibly even more basic than those measured by the psychological needs test. And age, income, family size, politics, and religion are typical demographic variables not necessarily associated with any specific behavior patterns.

Another group of objective factors that might be predictive of brand was purposely omitted. These are details of the car, such as size, color, horsepower. The obvious reason for deleting these variables is that, if they were known, the brand would be too. To use these would be to mix "dependent" and "independent" variables. Prediction would then be little more than a mathematical exercise. In this study, for example, the number of cylinders would have been an excellent, but trivial, discriminator between Ford and Chevrolet. Over twice as many six-cylinder Chevrolets as Fords occurred in the sample.

Many of the demographic variables are qualitative, and scaling them raises important problems. These problems were mainly bypassed by treating the qualitative variables in dichotomous fashion: respondents were placed in one class or another.

Religion and politics presented special problems because there is no obvious ordering of the categories. The method of dummy variables was used.²⁸ Religion was treated as two variables: Protestant or not, Catholic or not. Each of these variables is dichotomous. The remaining dummy variable, non-Christian or not, was not included explicitly; this avoids the statistical problem arising from a singular matrix.²⁹ The number of cases involved in the omitted dummy variable was small, less than 8 per cent of the sample. Implicitly, of course, the variable is included: "Non-Protestant and "non-Catholic" jointly define "non-Christian."

In like fashion, politics was split into two variables. Republican or not and Democrat or not. The third dummy variable, Republican or Democrat versus all others, was deleted as before. The use of these dummy variables increased the total objective variables to fourteen.

The intercorrelations of these variables were examined. For the fourteen variables, including the dummy variables, ninety-one comparisons are necessary. With the exception of the dummy variables, most of these intercorrelations are low. With the dummy variables high, negative intercorrelations were found, as expected. Catholic or not and Protestant or not have an intercorrelation of $-.838$; Republican or not and Democrat or not, $-.460$.

Outside of these, the highest intercorrelation is between age and more

²⁸ Daniel B. Suits, "Use of Dummy Variables in Regression Equations," *Journal of the American Statistical Association*, Vol. LII (1957), pp. 548-51.

²⁹ In regression or linear discriminant function solutions the normal equations can be solved by placing restrictions upon the equations—setting the constant term of the equation equal to zero. The common solution, however, is simply to drop one of the dummy variables (*ibid.*).

than five years with the same firm. The relationship is $+0.401$. The second highest non-dummy variable correlation is the relationship between Catholicism and having three or more children. The correlation coefficient is $+0.378$. Catholicism is also highly correlated with frequent church attendance ($+0.334$).

Group Scores of the Objective Variables

The average scores of Ford and Chevrolet owners are shown in Table 6. Also the scoring range and the differences between group means

TABLE 6
AVERAGE SCORES OF DEMOGRAPHIC VARIABLES FOR FORD AND CHEVROLET OWNERS

<i>Variable</i>	<i>Scoring Range</i>	<i>Ford</i> (<i>N</i> = 72)	<i>Chevrolet</i> (<i>N</i> = 74)	<i>Difference</i>
Age of car.....1 (1958) —4 (1955)		2.625	3.014	-0.389*
Over 10,000 miles per year....1 (over) —0 (under)		0.722	0.622	+0.100
Shopped more than one dealer.....1 (yes) —0 (no)		0.750	0.716	+0.034
Owner smokes.....1 (yes) —0 (no)		0.778	0.608	+0.170†
Own—rent.....1 (Rent) —0 (Own)		0.417	0.554	-0.137‡
Three or more children at home.....1 (yes) —0 (no)		0.444	0.311	+0.133‡
Catholic or not.....1 (yes) —0 (no)		0.319	0.230	+0.089
Protestant or not.....1 (yes) —0 (no)		0.639	0.662	-0.023
Attend church more than once a month.....1 (no) —0 (yes)		0.375	0.460	-0.085
Republican or not.....1 (yes) —0 (no)		0.444	0.378	+0.066
Democrat or not.....1 (yes) —0 (no)		0.181	0.284	-0.103
Age.....1 (19) —9 (54)		5.333	5.351	-0.018
Five or more years with same firm.....1 (yes) —0 (no)		0.625	0.473	+0.152‡
Income (mid-points).....1 (\$3,750)—6 (\$16,250)		3.194	3.068	+0.126

* Significant at 2 percent level (two-tailed test).

† Significant at 5 percent level (two-tailed test).

‡ Significant at 10 percent level (two-tailed test).

are given. There were no hypotheses concerning the direction of the differences, and a two-tailed test was used to compare group means.

Nine of the variables show no significant differences between means. Age, income, religion, and politics are among these. These are among the variables most commonly used in marketing research to describe specific brand markets. In addition to these, usage over 10,000 miles per year and "shopping" showed no differences between the two groups.

The most significant difference between the groups was shown in the age of the car owned. Fords were newer. Thirty-five of the seventy-two Fords were of 1957 or 1958 model compared to twenty-one out of seventy-four Chevrolets. This reflects the different popularity of the brands in different years.

In the universe sampled, Ford accounted for 53 percent of the owners, Chevrolet for 47 percent. The difference cannot be accounted for by

the popularity of various body styles alleged to be popular in suburbia. For some years Ford has claimed sales superiority in convertibles and station wagons, yet in the sample there were nineteen Fords of these models and sixteen for Chevrolet; in other models the distributions were also similar.

The two next largest differences were smoking and working for the same firm for five or more years. The Ford group contains more men who smoke and more men who have stayed with the same company for five or more years. Although these variables are objective in nature, they suggest differences in personality possibly more basic than those discussed earlier.

The two other variables showing significant differences between Ford and Chevrolet owners are homeownership and three or more children living at home. In each of these cases Ford owners show a higher percentage than Chevrolet owners. The intercorrelation of these two variables is only +.219. Both of these suggest strong family life and are not exactly what one would picture a Ford owner to be from the motivation research findings previously mentioned.

For all fourteen of these demographic variables the distributions of both groups overlap substantially. Although five of the fourteen group means are significantly different, this overlap reduces the chance for successful discrimination by any one variable.

Linear Discriminant Function

Weights of the Variables. The weights of the fourteen variables for the linear discriminant function are shown in Table 7. The group's means and the mean of the $\bar{Y}$ for each group are also given.

TABLE 7
LINEAR DISCRIMINANT FUNCTION OF DEMOGRAPHIC VARIABLES FOR FORD
AND CHEVROLET OWNERS

Variable	Discriminant Function Weight	Group Means	
		Ford (N = 72)	Chevrolet (N = 74)
Age of car.	+1.0000 X_1	2.6250	3.0140
Used over 10,000 miles per year.	-1.0480 X_2	0.7222	0.6216
Shopped before buying.	-0.1204 X_3	0.7500	0.7162
Owner smokes.	-2.1629 X_4	0.7778	0.6081
Homeowner—renter.	+1.0189 X_5	0.4167	0.5541
Three or more children at home.	-0.8388 X_6	0.4444	0.3108
Catholic or not.	-3.4376 X_7	0.3194	0.2297
Protestant or not.	-2.8371 X_8	0.6389	0.6622
Attend church more than once a month.	+0.2189 X_9	0.3750	0.4595
Republican or not.	+0.3198 X_{10}	0.4444	0.3784
Democrat or not.	+1.7266 X_{11}	0.1806	0.2838
Age.	-0.1304 X_{12}	5.3330	5.3510
Five or more years with same firm.	-1.0576 X_{13}	0.6250	0.4730
Income.	+0.2482 X_{14}	3.1940	3.0680
$\bar{Y}_i = \Sigma \text{weights times group means.}$		-2.7909	-1.1283

Predictive Ability. In terms of the number of cases classified correctly, this discriminant function is slightly better than the one using the psychological need variables. It misclassified only 30.1 percent of the cases, compared to 37.1 percent for the latter. This still leaves much to be desired in terms of predictive ability.³⁰ The number of cases classified by brand are shown in Table 8.

TABLE 8
CLASSIFICATION OF FORD AND CHEVROLET OWNERS
BY LINEAR DISCRIMINANT FUNCTION OF DEMOGRAPHIC VARIABLES

Brand	Classified Correctly	Misclassified	Total
Ford.....	52	20	72
Chevrolet.....	50	24	74
Total.....	102 (69.9%)	44 (30.1%)	146 (100%)

The multiple correlation coefficient (R) of this discriminant function is .3819, and its square (R^2) is .1458. Analysis of variance shows this equation to be statistically significant between the 10 and 5 percent levels, the same as the psychological need equation. This analysis of variance is shown in Table 9.

TABLE 9
ANALYSIS OF VARIANCE OF LINEAR DISCRIMINANT FUNCTION
OF DEMOGRAPHIC VARIABLES

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Discriminant function...	14	0.1458(R^2)	0.010414
Remainder.....	131	0.8542(1 - R^2)	0.006521
Total.....	145	1.000	
$F = 1.597$			

Other Objective Factors

In addition to the demographic variables used in the linear discriminant function, several other factors were investigated. In terms of education the two groups were very similar. The distributions by brand of highest grade level completed are almost identical.

Stock ownership in the automobile companies was very rare and not associated with the brand owned. Two Ford owners held Ford Motor Company stock and one owned General Motors stock. One Chevrolet owner had Ford stock; none had General Motors.

³⁰ The expected proportion of misclassification for this discriminant function is 0.355. For future samples this would be an understatement (see n. 22).

Twenty-four Ford owners and twenty-six Chevrolet owners had high fidelity (HiFi) phonographs in their homes. None owned color television sets.

Newspaper reading habits of the two groups were also extremely similar. Sixty to 65 percent of each group of owners read the *Chicago Tribune* and/or *Daily News*. About 30 percent read the *Sun-Times* and 10 percent the *American*.

Conclusion

The linear discriminant function of demographic variables is not a sufficiently powerful predictor to be of much practical use. It does a somewhat better job of classification than the function based upon psychological need scores. Its results, however, are not enough better to favor strongly the traditional research mode over motivation research. Both point more to the similarity of Ford and Chevrolet owners than to any means of discrimination between them. Analysis of several other objective factors also leads to the same conclusion.

IV. COMBINED ANALYSIS: PSYCHOLOGICAL AND OBJECTIVE FACTORS

An Eclectic Approach

The central problem of this paper dictated that the predictive abilities of the two kinds of data be treated separately. However, with the data available, one would not normally restrict the analysis to these separate and distinct comparisons. It is highly possible that some combination of demographic and psychological variables would be better than either alone. To test this, a linear discriminant function was computed using as independent variables those of each kind which showed the greatest differences between Ford and Chevrolet owners.

Selecting the independent variables for this analysis upon the basis of the earlier investigations should increase the probability that prediction will improve. Statisticians know well that in correlation and regression problems any competent statistician can produce highly significant results, given enough time and data to pick and choose, rejecting what does not work and retaining what seems promising. Testing hypotheses upon the data from which they were derived is statistically unsound.³¹ Therefore, the following analysis should be viewed with these limitations in mind.

Selection of Variables. The variables were selected upon the basis of their comparative significance in the previous discriminant functions.³²

³¹ W. A. Wallis and H. V. Roberts, *Statistics: A New Approach* (Glencoe, Ill.: Free Press of Glencoe, Inc., 1956), p. 405.

³² The variables selected for each type of data are those with the highest *F* ratios of their partial regression coefficients. This was possible because the original discriminant function was computed from the multiple regression model.

Six demographic variables and five psychological needs were chosen. The demographic variables are as follows: X_1 , smoking; X_2 , homeownership; X_3 , three or more children at home; X_4 , X_5 , religion; X_6 , X_7 , politics; and X_8 , five or more years with the same company. As before, religion and politics were each split into two through the use of dummy variables. Thus in the discriminant function these six actually became eight. The five psychological needs that looked most promising are the following: X_9 , deference; X_{10} , exhibition; X_{11} , autonomy; X_{12} , affiliation and X_{13} , dominance.

Linear Discriminant Function

Weights of the Variables. Table 10 gives the weights for each of these thirteen variables in the combined analysis. Group means and the $\bar{Y}$ for each group are also shown.

TABLE 10
LINEAR DISCRIMINANT FUNCTION COMBINING OBJECTIVE AND PSYCHOLOGICAL
NEED VARIABLES FOR FORD AND CHEVROLET OWNERS

Variable	Discriminant Function Weight	Group Means*	
		Ford ($N = 71$)	Chevrolet ($N = 69$)
Owner smokes.....	-1.0000 X_1	0.778	0.608
Homeownership.....	+0.6631 X_2	0.417	0.554
Three or more children at home.....	-0.4919 X_3	0.444	0.311
Catholic or not.....	-1.1735 X_4	0.319	0.230
Protestant or not.....	-0.8286 X_5	0.639	0.662
Republican or not.....	+0.1019 X_6	0.444	0.378
Democrat or not.....	+0.7282 X_7	0.181	0.284
Five or more years with same firm.....	-0.1884 X_8	0.625	0.473
Deference.....	+0.0137 X_9	9.470	9.770
Exhibition.....	-0.0634 X_{10}	10.060	9.300
Autonomy.....	+0.1376 X_{11}	7.860	8.800
Affiliation.....	+0.1616 X_{12}	10.140	11.090
Dominance.....	-0.1151 X_{13}	13.690	12.410
$\bar{Y}_i = \Sigma \text{weights times group means}.....$		-0.9280	+0.0630

* See Table 1 for personality need scores; Table 6 for objective variable scores.

Predictive Ability. Although "loaded" to produce favorable results, this discriminant function does not show better predictive ability than the one based on demographic factors alone. It is only slightly better than the one based upon psychological needs. It misclassified 51 out of the 140 cases from which it was developed; 36.4 percent were assigned to the wrong brand.³³ By comparison, the demographic factor discriminant function misclassified 30.1 percent of the cases and the psychological need discriminant function, 37.1 percent. Classification by brand by the combined factor discriminant function is shown in Table 11.

This apparent lack of improvement of the combined variable dis-

³³ The expected proportion of misclassified cases for this combined discriminant function is 0.340. As before, this understates the real problem (see n. 22).

TABLE 11

CLASSIFICATION OF FORD AND CHEVROLET OWNERS BY LINEAR
DISCRIMINANT FUNCTION USING BOTH PSYCHOLOGICAL AND
DEMOGRAPHIC VARIABLES

<i>Brand</i>	<i>Classified Correctly</i>	<i>Misclassified</i>	<i>Total</i>
Ford.....	46	25	71
Chevrolet.....	43	26	69
Total.....	89 (63.6%)	51 (36.4%)	140 (100%)

criminant function over the demographic factor function may be due to not including the age of automobile owned as a variable in the combined analysis. This variable showed the greatest separation of the group means among the demographic variables. Ideally, comparisons would be made only for specific years, and this kind of "semidependent" variable would not enter in. Sample size prohibited this, however, in this study.

The multiple correlation coefficient of the combined variable discriminant function is .3991, the highest of the three functions computed. Analysis of variance shows the equation to be statistically significant at the 5 per cent level. This is shown in Table 12.

TABLE 12

ANALYSIS OF VARIANCE OF LINEAR DISCRIMINANT FUNCTION
USING DEMOGRAPHIC AND PSYCHOLOGICAL NEED VARIABLES

<i>Source of Variation</i>	<i>Degrees of Free- dom</i>	<i>Sum of Squares</i>	<i>Mean Square</i>
Discriminant function . . .	13	0.1593(R^2)	0.012254
Remainder.....	126	0.8407($1 - R^2$)	0.006672
Total.....	139	1.0000	
$F = 1.837$			

Conclusion

The combination of the most predictive variables from the two earlier analyses did not produce an effective method for distinguishing between Ford and Chevrolet owners. The linear discriminant function combining these variables is statistically significant at a higher level than the previous two, but its discriminatory ability is still low. This combined analysis does not show any significant superiority over the earlier ones, nor does it point toward some combination of the variables as explaining the choice between Ford and Chevrolet.

V. OTHER ASPECTS OF PURCHASE BEHAVIOR

Although subsidiary to the main investigation, interview data were collected specifically for analysis of other areas of behavior commonly

associated with consumer products. These areas are (1) the image or stereotype that a particular brand may have in people's minds and (2) differences between owners who are loyal to one brand as opposed to those who switch brands. Each of these areas was examined for differences between Ford and Chevrolet owners and for clues to their purchase motives.

Brand Stereotypes

Brand Images. A brand's image consists of all the things associated with the product or perceived about it. It is a total personality extending beyond its physical qualities. And it is this image that people are thought to purchase rather than any reality. Also an image is consistent among both users and nonusers; it is commonly held views of brands that attract some customers and repel others. The image specifies what kinds of people or for what uses the particular brand is best suited.

A brand image is the result of three distinct forces. First, the product itself by nature of its physical makeup and design may be better for some uses or kinds of people than other competing products. Second, the manufacturer through his advertising tries to create the impression that his brand is best for certain people or uses. For example, a certain automobile has recently been advertised as being particularly appropriate for doctors, as it is a very dependable car. And, third, people associate a brand with the type or classes of people they observe using it. In some undefined and unspecified pattern, these elements contribute to the brand's personality. Advertising stresses the importance of the second of these factors. Current market research activities are often oriented toward the third.

Research on Automobiles. A study by Munn indicated that consumers perceive significant quality differences between brands of automobiles.³⁴ He also found these images to be independent of age, education, or income of the consumer. More specifically, a study by the Bureau of Applied Social Research indicated that among the low-priced automobiles the Ford owner was perceived as being more youthful, more masculine, and of lower social class than owners of Chevrolets or Plymouths.³⁵

Assessment of Brand Images. To measure the brand images of Ford and Chevrolet respondents were given the following instructions and twenty-one brief descriptions of people:

We often think of cars especially suitable or unsuitable for different kinds of people, the way it seems odd for a very big man to drive around in a very

³⁴ Henry L. Munn, "An Exploratory Investigation of Brand Perceptions by Specified Classes of Consumers for Specified Classes of Consumer Goods" (unpublished Ph.D. dissertation, School of Business, University of Chicago, 1957), p. 37.

³⁵ Bernard Levenson *et al.*, "Social Stereotypes of Automobile Makes" (unpublished research report, Bureau of Applied Social Research, Columbia University, June, 1956), pp. 5-30.

small car. I'll describe some people for you and I'd like you to tell me if a Ford or Chevrolet would be a better car for each of them. Even if you think any car will do, name the one that comes closest, in your opinion, in any way.

1. He likes to solve difficult problems, always does his best.
2. He always gets advice from others before buying anything.
3. He's always telling jokes and using big words.
4. He does whatever he pleases, doesn't like to conform.
5. He is really loyal to his friends.
6. He likes to observe others and understand how they feel.
7. He wants to be the boss, to supervise others.
8. He likes to be punished when he's wrong.
9. He likes to travel, meet new people; enjoys change.
10. He thinks he is physically attractive to women.
11. An aggressive driver, always first away from the light.
12. He's an athlete—very masculine type.
13. In picking a job, he looks for security.
14. You can tell he's a college boy.
15. A person who is very self-confident.
16. He's a very cautious driver, never had a ticket.
17. The best car for a woman.
18. You can tell he's a successful man.
19. Loves his car, tinkers with it all the time.
20. A very dignified and reserved gentleman.
21. Everything he owns is the latest style.

Seventy Ford and seventy-four Chevrolet owners completed this portion of the interview.

The Images of Ford and Chevrolet. For the combined sample of Ford and Chevrolet owners there is agreement on fourteen of the twenty-one descriptions. That is, the combined group gave a majority for the same car on fourteen items of the twenty-one. The percentage of respondents agreeing that either a Ford or a Chevrolet was the more appropriate car ranged from a low of 59.03 per cent to a high of 77.78 per cent. All are statistically significant at the 5 per cent level or beyond.

However, looking at Ford and Chevrolet owners as separate groups rather than combined shows that in nine of the fourteen images there is not true consensus independent of the brand owned, that is, while each group gave a majority to the same car, the extent of the majority was statistically significant only five times. These nine are not true images by rigorous definition.

In only five of the images is there agreement statistically significant at the 5 percent level for each group separately as well as combined. In four of these five, the Ford owner is pictured as one who does not like to conform, is an aggressive driver, loves his automobile, or is a college boy. In the fifth, the Chevrolet owner is pictured as a very cautious driver. These five instances present the only clear-cut brand images discovered in this study. The percentages of consensus are shown in Table 13.

The other nine images which showed agreement for the combined

TABLE 13
BRAND IMAGES OF FORD AND CHEVROLET REPRESENTING
COMPLETE CONSENSUS

<i>Social Image</i>	<i>Percentage of Combined Sample (N = 144)</i>	<i>Percentage of Ford Owners (N = 70)</i>	<i>Percentage of Chevro- let Owners (N = 74)</i>
Ford owner:			
Does whatever he pleases	66.0*	68.6*	63.5*
An aggressive driver	77.8*	72.9*	82.4*
A college boy	67.4*	65.7*	68.9*
Loves his car	64.6*	61.4*	67.6*
Chevrolet owner:			
Very cautious driver	77.8*	64.3*	90.5*

* Reject null hypothesis at 5 percent level of significance (test figures carried to four decimal places).

group of owners presents two distinct pictures of the automobiles. Ford owners have views on Fords and Chevrolets that Chevrolet owners believe are unimportant. The reverse is also true.

Ford owners see their brand as best for an athlete and for someone always up to date. They picture the Chevrolet owner as always seeking advice from others. Chevrolet owners do not assign these images to either brand at a significant level. The percentage of Ford owners holding these views is shown in Table 14.

TABLE 14
PERCENTAGE OF FORD OWNERS HAVING IMAGES OF
FORDS AND CHEVROLETS WHICH CHEVROLET OWNERS
DO NOT HAVE

<i>Social Imagery</i>	<i>Percentage of Ford Owners (N = 70)</i>
Ford owners:	
An athlete	84.3*
Everything owned is latest style	68.6*
Chevrolet owners:	
Seeks advice from others	62.9*

* Reject null hypothesis at 5 percent level of significance (test figures carried to four decimal places).

Similarly, Chevrolet owners see the Chevrolet as best for a woman, for a dignified gentleman, or a person desiring job security. They see the Ford owner as always telling jokes, wanting to be boss, and thinking of himself as physically attractive. Ford owners see no differences between the brands with respect to these six images. The percentage of Chevrolet owners holding these images is shown in Table 15.

Brand Images Projected into Own Car by Both Groups. The other seven brand images presented were attributed by each group to their

TABLE 15
PERCENTAGE OF CHEVROLET OWNERS HAVING IMAGES OF
FORDS AND CHEVROLETS WHICH FORD OWNERS
DO NOT HAVE

<i>Social Imagery</i>	<i>Percentage of Chevrolet Owners (N = 74)</i>
Chevrolet owners:	
Desires job security.....	86.5*
Dignified gentleman.....	83.8*
Best for a woman.....	87.8*
Ford owners:	
Always telling jokes.....	71.6*
Wants to be boss.....	63.5*
Physically attractive.....	73.0*

* Reject null hypothesis at 5 percent level of significance (test figures carried to four decimal places).

own brand. They represent positive values to both groups. Both Ford and Chevrolet owners believe their brand is best for the person who:

1. Likes to solve difficult problems.
2. Is loyal to his friends.
3. Likes to observe and understand other people.
4. Likes to be punished when wrong.
5. Likes to travel and enjoys change.
6. Is very self-confident.
7. Is a successful man.

With the exception of the wish to be punished, these are socially desirable traits. The percentages assigning this latter image to their own brand are not significantly different from chance at the 5 percent level. The percentages of owners assigning the seven images to their own brand are shown in Table 16.

TABLE 16
PERCENTAGE OF FORD AND CHEVROLET OWNERS ATTRIBUTING BRAND
IMAGE TO THEIR OWN CAR

<i>Social Imagery</i>	<i>Percentage of Ford Owners Answering Ford (N = 70)</i>	<i>Percentage of Chevrolet Owners An- swering Chev- rolet (N = 74)</i>
Likes to solve difficult problems.....	65.7*	60.8*
Loyal to friends.....	74.3*	82.4*
Likes to observe others.....	60.0*	78.4*
Likes to be punished.....	58.6	58.1
Likes to travel.....	70.0*	67.6*
Self-confident.....	75.7*	66.2*
A successful man.....	62.9*	59.5

* Reject null hypothesis at 5 percent level of significance (test based on figures to four decimal places).

Brand Images and Personality Needs. This weakness of brand images for Ford and Chevrolet suggests that what is commonly thought of as a brand image is somehow a function of the individual's particular personality. The first eleven brand-image descriptions used in this study were taken directly from the explanation of the personality needs measured by the Personal Preference Schedule.³⁶ The brand-image descriptions were made parallel to the personality needs, so that the relationship between image and personality could be analyzed. The need-image pairings are as follows:

1. Achievement—He likes to solve difficult problems, always does his best.
2. Deference—He always gets advice from others before buying anything.
3. Exhibition—He's always telling jokes and using big words.
4. Autonomy—He does whatever he pleases, doesn't like to conform.
5. Affiliation—He is really loyal to his friends.
6. Intraception—He likes to observe others and understand how they feel.
7. Dominance—He wants to be the boss, to supervise others.
8. Abasement—He likes to be punished when he's wrong.
9. Change—He likes to travel, meet new people; enjoys change.
10. Sex—He thinks he is physically attractive to women.
11. Aggression—An aggressive driver, always first away from the light.

Analysis of these pairs shows that whatever need an individual indicated was most important to himself the corresponding image description was assigned to the car he owned far oftener than one would expect from chance alone—65.1 percent of the time for the total sample. Thus, in almost two-thirds of the cases, the individual projects his greatest need into the brand he happens to have. In this sense, then, automobiles are extensions of the owner's personality. That is, owners use their automobiles to satisfy certain personality needs that are important to them. They believe the car fits their personality and ascribe it to people of similar needs. However, the lack of brand discrimination by the personality need variables also precludes discrimination on the basis of these brand-image projections. While individuals tend to project their own needs into their cars, the distribution of needs is similar for owners of each make.

Considering all the needs ranked in the top half of each individual's scale, 61.2 percent of these are attributed to the brand owned. Slightly more Chevrolet owners than Ford owners project their top need into their car, but for the top five needs the pattern is reversed. Table 17 shows the percentages for each brand as well as the total sample.

Conclusion. In view of all the marketing literature of recent years, the brand images found were much less focused than one would expect. For Chevrolet, only one strong image appeared, and for Ford, only four. This suggests that the exploitation of these images will be difficult. These are the two most popular and best-advertised automobiles. This

³⁶ Edwards, *op. cit.*, p. 14.

TABLE 17
PERCENTAGES OF INDIVIDUALS PROJECTING PERSONALITY NEEDS
INTO CAR THEY OWN

	<i>Ford Owners (Percent)</i>	<i>Chevrolet Owners (Percent)</i>	<i>Both Combined (Percent)</i>
Single greatest need projected into own car.....	62.1	68.2	65.1
Five greatest needs projected into own car.....	64.2	58.0	61.2

finding could be due to the choice of image questions used in this study, but the factors selected were chosen after a thorough study of the current literature. Open-ended question techniques might have found more.

The relationship between the owner's personality needs and the brand images, however, suggests that brand images are not the independent phenomenon they are usually thought to be. The needs that a person values greatest he tends to assign to whatever brand he happens to have.

Brand Loyalty

Stability of Brand Preference. Customers who purchase the same brand repeatedly are often thought to be different from those who switch brands. For purposes of this study, owners whose previous car was the same as their present one were classified as brand-loyal. Thirty Ford owners and forty-three Chevrolet owners fitted this category. The percentages of loyal owners are 41.81 for Ford and 58.11 for Chevrolet. This reflects, again, increasing popularity for Ford, at least in part. Non-loyal segments include only one Ford and one Chevrolet owner whose present car was their first.

The differences between loyal and non-loyal owners are apparent in both their shopping habits and their future automobile plans. Less than two-fifths of the loyal owners shopped other brands before buying, compared to three-fifths of the non-loyal owners. Also over two-thirds of the loyal owners plan to remain loyal compared to a little over two-fifths of the non-loyal owners. By brand, the percentages vary somewhat, but the trend is constant. Table 18 shows these percentages by brand and for the combined group.

Not only were there more loyal Chevrolet owners in the sample, but also the evidence in Table 18 points to stronger loyalty. Fewer loyal Chevrolet owners shop other makes than do Ford owners, and more loyal Chevrolet owners plan to stick to Chevrolet in the future.

However, in this study the reason for classifying owners as loyal or non-loyal is not to try to explain this particular kind of behavior. Rather, it is to see whether, within a brand, the loyal and non-loyal owners represent two distinct types of people. If this is the case, lumping all Ford

TABLE 18

PERCENTAGE OF LOYAL AND NONLOYAL FORD AND CHEVROLET OWNERS WHO SHOPPED OTHER MAKES AND WHO PLAN TO BUY SAME CAR AGAIN

	<i>Ford</i>		<i>Chevrolet</i>		<i>Combined</i>	
	<i>Loyal</i> (<i>N</i> = 30)	<i>Nonloyal</i> (<i>N</i> = 42)	<i>Loyal</i> (<i>N</i> = 43)	<i>Nonloyal</i> (<i>N</i> = 31)	<i>Loyal</i> (<i>N</i> = 73)	<i>Nonloyal</i> (<i>N</i> = 73)
Percentage who shopped other makes before buying.....	46.7	54.8	32.6	67.7	38.4	60.3
Percentage planning to buy same brand again.....	60.0	40.5	72.1	41.9	67.1	42.5

owners into one group and all Chevrolet owners into the other could have caused confounding of the previous results, i.e., discrimination might have been possible, had loyal and non-loyal owners of each brand been treated separately. For each brand the loyal and non-loyal owners were compared with respect to the personality needs and the demographic variables which showed the greatest separation of the brands.

Personality Needs. The four personality needs which showed greatest differences between the Ford and Chevrolet owners are affiliation, autonomy, dominance, and exhibition. Comparison of the rank-order scores for these needs out of the eleven measured by loyal and non-loyal owners shows very little difference. For Ford, the two groups—loyal and non-loyal—rank the four needs almost identically. Loyal Chevrolet owners ranked dominance, exhibition, and autonomy slightly lower than non-loyal owners and placed affiliation higher on this scale.

If the loyal and non-loyal owners of each brand have widely differing personality need structures there is no indication of it here. The rankings by brand are shown in Table 19.

TABLE 19

RANK-ORDER SCORES OF SELECTED PERSONALITY NEEDS FOR LOYAL AND NONLOYAL FORD AND CHEVROLET OWNERS

<i>Personality</i> <i>Need</i>	<i>Ford</i>		<i>Chevrolet</i>	
	<i>Loyal</i> (<i>N</i> = 29)	<i>Nonloyal</i> (<i>N</i> = 42)	<i>Loyal</i> (<i>N</i> = 39)	<i>Nonloyal</i> (<i>N</i> = 30)
Dominance.....	3.2	3.1	4.7	3.5
Affiliation.....	5.7	5.5	4.7	5.6
Exhibition.....	6.2	6.1	6.8	6.5
Autonomy.....	7.9	7.7	7.4	6.9

Demographic Variables. Similarly, the loyal and non-loyal owners of each brand were compared with respect to the four demographic variables that showed the greatest separation of the brands. The variables

are homeownership, three or more children at home, smoking, and working for the same company for five or more years. Table 20 shows the percentages by brand for loyal and non-loyal owners.

On six of the eight comparisons, loyal and non-loyal owners are most similar. As before, loyal and non-loyal Ford owners are more alike than the comparable Chevrolet owners. Non-loyal Chevrolet owners own

TABLE 20

COMPARISON OF LOYAL AND NONLOYAL FORD AND CHEVROLET OWNERS BY SELECTED
DEMOGRAPHIC VARIABLES

<i>Demographic Variables</i>	<i>Ford</i>		<i>Chevrolet</i>	
	<i>Loyal</i> (<i>N</i> = 30)	<i>Nonloyal</i> (<i>N</i> = 42)	<i>Loyal</i> (<i>N</i> = 43)	<i>Nonloyal</i> (<i>N</i> = 31)
Percentage homeowners.....	53.3	59.5	39.5	51.6
Percentage families with three or more children... .	46.7	42.8	30.2	32.3
Percentage smokers.....	76.7	78.6	58.1	64.5
Percentage working for same firm for five or more years.....	60.0	59.5	41.9	54.8

homes more frequently than do their loyal counterparts, and more of them have worked for the same company for five or more years. However, even the largest difference shown is not significant at a 5 percent level.

The loyal and non-loyal owners of each brand appear to be essentially the same with respect to the variables used in this study. Pooling loyal and non-loyal owners into the same group does not wash out differences that might otherwise be important.

VI. SUMMARY AND CONCLUSIONS

Although respondents in this survey spanned wide ranges for most of the psychological and objective variables measured, none of these variables was systematically related to the brand of car owned for the two brands which constitute almost half the automobile market. The variables used here do not allow for further segmentation of this market.³⁷

Two major limitations are recognized. The success of Ford and Chevrolet over the years attests to their appeal to many kinds of people. A comparison of owners of similarly priced but less popular automobiles, like the Rambler, with owners of either Ford or Chevrolet might show greater discrimination with the variables used. The writer's experiences with this study, however, cause him to be skeptical.

³⁷ See also Gladstone Bonnick, "The Condition of Chevrolet and Ford Cars Owned by Negroes in Chicago" (unpublished term paper, Graduate School of Business, University of Chicago, June, 1959). Bonnick found no significant differences between Fords and Chevrolets of 1956 through 1959 models in need of obvious repairs. Neither group seemed to keep its cars in better condition.

Second, the Park Forest universe from which the sample was drawn is in no way to be construed as representative of the entire automobile market. On the other hand, restricting this study to a relatively homogeneous group in certain respects, such as age and income, makes it possible to examine other variables, especially psychological ones, with greater precision. The problem of distinguishing between owners (and prospective owners) of highly competitive brands in a homogeneous market area closely parallels problems facing the manufacturer and his advertising agency, unless it can be shown that Park Forest is wholly atypical, *in respects here studied*, of the rest of the country.

If an individual's personality and/or his demographic characteristics can be used to predict the choice between a Ford and Chevrolet, different measures and techniques must be found. The variables included in this study do not explain brand choice, and their discriminatory ability is much less than previous research has indicated. It seems to make little difference to a large percentage of car owners with widely varying psychological and other characteristics whether they own a Ford or a Chevrolet. These makes appear substitutable in many areas besides price.

This study does not point to the clear-cut superiority of either research mode. Over-all, the objective factors did a somewhat better job of discrimination but still an unsatisfactory one. Table 21 compares the

TABLE 21
COMPARISON OF LINEAR DISCRIMINANT FUNCTIONS DESCRIBING FORD AND CHEVROLET OWNERS

<i>Class of Independent Variables</i>	<i>No. Vari- ables Em- ployed</i>	<i>Multiple Cor- relation Coeffi- cient of Dis- criminant Function</i>	<i>F Ratio of Discriminant Function</i>	<i>Percentage Sample Mis- classified by This Equa- tion</i>
Psychological needs	10	0.3353	1.634	37.1
Demographic factors	14	0.3819	1.597	30.1
Selected combination of both	13	0.3991	1.837*	36.4

* Significant at the 5 percent level.

three discriminant functions computed in this study. The design of the study placed some broad limitations upon the objective factors that would not be expected to apply to the psychological needs. Many motivation researchers have claimed that it is the lack of discrimination by objective variables that produces the need for their wares. This study does not bear them out.

From the standpoint of marketing strategy, this study highlights the difficulties involved in segregating the customers of one brand from those of a similar and competing one. Popular brands appeal to differ-

ent kinds of people for many different reasons. What motivates these customers is not readily apparent.

By definition, a brand image cannot attract opposite kinds of people, i.e., an image must be consistent. Even with some product variation, it may be impossible for a manufacturer to create several distinct images. Within a brand family there are carry-overs from their common heritage. Also to try to create several images rather than a single one may cause a "scattering" of promotional effort and leave one open to the inroads of competitors.

This study suggests that many of the commonly held assumptions in marketing about brand images are either wrong or misleading. The evidence points neither to strong images attracting definite kinds of people nor to the use of automobiles for satisfying deep inner needs in symbolic terms.

In promoting a brand it would appear safest to be somewhat ambiguous for both personality and objective variables. People of all kinds are customers, and creation of too strong an image in certain personality terms may narrow one's market unnecessarily. If the image is ambiguous, there is a tendency for customers to read into it what they want. The things they value highly they attribute to their brand.

The implications of purchase motivations in this research may be viewed in two ways, not necessarily mutually exclusive. One is that people choose automobiles on the basis of obvious "rational" factors: lowest price, comparison of mechanical features, operating performance, etc. Second, brand choice may depend upon small things, peculiar to the individual, not usually measured in marketing research. These latter motivations can be characterized as idiosyncratic.

TECHNICAL APPENDIX

The purpose of this appendix is to give details and methodology that were a necessary part of this study, although not essential to the text or the results as presented.

The Park Forest Universe

Park Forest is a post-World War II suburban community approximately thirty miles south of the Chicago Loop. It was developed by one builder from previous farm lands. It has received wide attention from sociologists.

A suburban community rather than part of Chicago proper was selected for this study for three major reasons:

1. Sociologists continually point to the suburbs as the pace-setters for American life. David Riesman, for example, recently said, "Suburbs in the last dozen years have become a symbol of the American way of life."³⁸ If trends exist in these areas, then they are of greater importance.

2. Population in the suburban areas has grown much faster than the rest of the country. The Census Bureau reports that from 1950 to 1956 the out-

³⁸David Riesman, "The American Future," speech given at the University of Chicago, February 3, 1958; reported in the *Chicago Maroon*, February 7, 1958, p. 3.

lying parts of standard metropolitan areas grew about six times as rapidly as did the central cities.³⁹

3. It was previously known that results of the psychological test used are influenced by demographic variables.⁴⁰ The selection of a homogeneous suburb reduces the ranges (and effects) of these variables.

The universe was restricted to owners of 1955 or later models and to males with only one car registered in their names. The universe list consisted of 869 Ford owners and 770 Chevrolet owners who had purchased a \$5.00 local vehicle-tax stamp. As local enforcement of this tax was alleged to be vigorous, it was believed that the listing would be accurate. However, when it came actually to interviewing those selected from this list, it was found to be almost 20 percent in error. Of the first two hundred names randomly selected from the list, thirty-eight were in error. Eighteen had moved since purchasing the tax stamp, and twenty owned a different make of car. In the sampling, these thirty-eight cases were treated as if they were not on the list to begin with, i.e., they were ignored, and the next name selected was used.

A simple random sample was selected by assigning each person in the universe a four-digit random number. The numbers were taken from the RAND Corporation, *A Million Random Digits with 100,000 Normal Deviates* (Glencoe, Ill.: Free Press of Glencoe, Inc., 1955). The names were then arrayed by these random numbers and the first one hundred of each brand chosen as the sample. The errors in the universe list required more than two hundred names to be drawn before the sample was completed.

Data Collection

Interviewing. Thirteen different women were employed as interviewers. All were married women with previous interviewing experience, and all worked only part time. They were regular employees of an interviewing service. As the respondents were all employed men, the interviewing was restricted to Saturdays, Sundays, and the early evening hours on other days. Interviewing time ranged from forty to one hundred minutes, with the average being about sixty-five minutes.

A careful verification was made of each interviewer's work by both telephone and postal card. Although personal interview with the male members of the family was specified, the verification form asked who in the family was interviewed, how they were interviewed, if they personally filled out a long list of paired statements (the psychological schedule), and if the interviewer was courteous. In addition, the telephone verifications repeated several items from the questionnaire to check upon the accuracy of the reporting. This verification uncovered a substantial amount of cheating by the interviewers. Four of the thirteen women employed were found to have cheated on parts of a total of thirty-five interviews. These thirty-five contaminated interviews were discarded, and no other work was accepted from these women.

Other Response Factors. Before the field work was completed, a total of two hundred and sixty-five names was used in the sample. Besides the thirty-eight errors on the universe list and the thirty-five partially faked interviews, an additional forty-six could not be reached after at least four follow-ups. There were twenty-one refusals, and an additional twenty-five respondents could not be located. All of these missing cases are shown in Table 22.

³⁹ U.S. Bureau of the Census, *Current Population Reports* (Ser. P-20, No. 71 [December 7, 1956]), p. 1.

⁴⁰ Arthur Koponen, "The Influence of Demographic Factors on Responses to the Edwards Personal Preference Scale" (unpublished Ph.D. dissertation, Columbia University, 1957).

TABLE 22

UNOBTAINABLE AND UNUSABLE INTERVIEWS WITH FORD AND CHEVROLET OWNERS IN PARK FOREST

Source of Difficulty	Total	Ford	Chevrolet
Errors in universe listing.....	38	22	16
Interviewer cheating.....	35	20	15
Respondent refused to be interviewed.....	21	7	14
Respondent not at home.....	25	14	11
Total.....	119	63	56

Thus, to secure 146 useful interviews, 265 randomly selected names were taken from the universe. This is a completion percentage of 55.1. Even eliminating these interviews unsecured because of errors in universe listing and those lost by cheating, only 76.0 percent of the selected names could be interviewed.

The Psychological Test

Question of Normalizing the Test Scores. Often psychological test scores are normalized around a common mean. In his manual, Edwards gives normalized scores with a mean of 50 and a standard deviation of 10 for each of the needs.⁴¹ In this study only the raw need scores from the psychological schedule were used. Normalizing each of the groups with a common mean and standard deviation performs no obviously useful function other than helping to achieve one of the assumptions of the discriminant function model. But, even here, normalizing each needs distribution separately is no guaranty that the joint distribution would be multivariate normal.

Almost half a century ago, statisticians debated the merits of assuming that observations arise from underlying normal continua. Karl Pearson and David Heron on one side claimed that the assumption is very often justified.⁴² G. Udny Yule vigorously and acrimoniously opposed them, arguing that the assumption is often artificial.⁴³ Today Yule's position is more generally accepted among statisticians, though Pearson has won the field in psychology. The writer, being more statistician than psychologist, chose to follow Yule's reasoning and hence did not normalize the test scores.

Test Reliability. With the sample of Ford and Chevrolet owners it was not possible to make direct tests for reliability of the measures. However, with his normative group, Edwards was satisfied that answering patterns were not random or haphazard.⁴⁴ Also, for a group of eighty-nine students, he found that test-retest correlations after a one-week period ranged from +.74 to +.88.⁴⁵

To see whether the test as taken by the Ford and Chevrolet owners was

⁴¹ Edwards, *op. cit.*, pp. 12-13.

⁴² Karl Pearson and David Heron, "On Theories of Association," *Biometrika*, Vol. IX (1913), pp. 159-315.

⁴³ G. Udny Yule, "On the Methods of Measuring Association between Two Attributes," *Journal of the Royal Statistical Society*, Vol. LXXV (1912), pp. 579-642; see also Leo A. Goodman and William H. Kruskal, "Measures of Association for Cross Classification," *Journal of the American Statistical Association*, Vol. XLIX (1954), pp. 735-36.

⁴⁴ Edwards, *op. cit.*, pp. 10-11.

⁴⁵ *Ibid.*, pp. 16-17.

operating in its usual way, simple rank-order correlations of scores on ten needs (those used in the discriminant function) were made, comparing the combined sample in this study with other published groups. With Edwards' normative group of 760 college men the rank correlation of needs is .903.⁴⁶ Similar comparison with 953 males living in market areas of over two million people and belonging to the J. Walter Thompson Company Consumer Panel showed a rank correlation of .806.⁴⁷ The rank correlation of nine of the needs with a group of forty hospitalized paranoid males was only .383.⁴⁸ Thus there is reason to believe that the test was functioning normally and that shortening it did not seriously affect the results, and even possibly that Park Foresters are not paranoid.

Linear Discriminant Function

The linear discriminant function was introduced by R. A. Fisher in the mid-1930's.⁴⁹ The statistical purpose of the function is to provide the maximum separation of two groups by maximizing the ratio of the difference between the specific means to the standard deviations within the groups. Fisher does not give any rationale for restricting his solution of this problem to linear equations, but Hodges shows that when the observations arise from normal populations and have the same covariance matrix, the function has certain optimum properties.⁵⁰

Fisher shows that the solution of the linear discriminant function is essentially the same as that for multiple regression when the dependent variable is dichotomous and can be assigned values separated by unity.⁵¹ The equations for the discriminant function and those for multiple regression differ only by a constant factor on the right-hand side. Also, Garrett has shown that the coefficient weights of multiple regression and the discriminant function are exactly proportional in the dichotomous case.⁵² Although in existence for over twenty years, the discriminant function has had little application to business problems. In a bibliography of over 250 uses of it listed by Hodges, only four are concerned with problems of the business world.⁵³ Although typically limited to comparison of two groups, the function has been expanded for use with three or more groups under certain conditions.⁵⁴

The computations for the linear discriminant functions used in this study were done on the UNIVAC I belonging to the University of Chicago Operations Analyses Laboratory. A multiple regression program was employed. Solution of the regression equation with 146 observations for each of fifteen independent variables took less than eight minutes computer running time. The

⁴⁶ *Ibid.*, p. 10. Reject the null hypothesis at better than the 0.01 level.

⁴⁷ Koponen, *op. cit.*, p. 27. Reject the null hypothesis at better than the 0.01 level.

⁴⁸ Mack Knutsen, "An Empirical Comparison of the Linear Discriminant Function and Multiple Regression Techniques in the Classifying Subjects into Three Categories" (unpublished Ph.D. dissertation, University of Washington, 1955), p. 26b. The null hypothesis would not be rejected at a 10 per cent level of significance.

⁴⁹ Fisher, *op. cit.*

⁵⁰ Joseph L. Hodges, Jr., *Discriminatory Analysis* (Randolph AFB, Texas: School of Aviation Medicine, USAF, 1955), chap. viii.

⁵¹ Fisher, *op. cit.*, pp. 184-85.

⁵² Henry E. Garrett, "The Discriminant Function and Its Uses in Psychology," *Psychometrika*, Vol. VIII (1943), pp. 65-79.

⁵³ Hodges, *op. cit.*, pp. 47-52.

⁵⁴ C. R. Rao, *Advanced Statistical Methods in Biometric Research* (New York: John Wiley & Sons, Inc., 1952), pp. 307-29.

savings of both time and money, to say nothing of the error possibilities, over hand computation were almost unbelievable. The increasing availability of electronic computers should make for greater use of discriminatory analyses in the future.

Psychologists Judging by Personality Needs

The eighteen psychologists picked only 70 correct out of 180 possible choices when asked to match the car brands with randomly selected personality need profiles. The distribution of correct choices and the statistical probabilities of these occurring by chance alone are shown in Table 23. The

TABLE 23
PROBABILITY OF OUTCOMES FOR CLASSIFICATION OF
FIVE FORD AND FIVE CHEVROLET OWNERS

<i>Number Correctly Placed in Own Group</i>	<i>Probability of Each Outcome under Null Hypothesis</i>	<i>Outcomes of Judging by 18 Psychologists</i>
10.....	1/252	
8.....	25/252	
6.....	100/252	1/18 (14/252)
4.....	100/252	15/18 (210/252)
2.....	25/252	2/18 (28/252)
0.....	1/252	

chi-square goodness-of-fit test for Table 23 shows non-random discrimination by the judges significant at the 1 percent level. But their selections were not positively correlated with the facts of brand ownership.



Probability Models for Brand Loyalty

THE RANGE OF STATISTICAL TECHNIQUES used in the study of brand choice with respect to frequently purchased food and household products is relatively great. Nevertheless, its treatment as a separate category is justified by the fact that it is one of the principle areas of marketing in which rigorous theoretical and empirical approaches have been brought together.

The Fourt and Woodlock article, "Early Prediction of Market Success for New Grocery Products," presents a simple probability model which describes the market penetration of a new product. It focuses upon the probability that a consumer will repeat the purchase of a given item. This is estimated by the empirical repeat ratio, or fraction of buyers who actually make the subsequent purchase. Prediction capability, as differentiated from the description of past behavior, depends upon the accurate assessment of the time shape of the penetration function. The authors state that their experience with a large number of new products has enabled them to determine the general shape of this curve, while observations on the early market behavior of a particular new product can be used to determine its constants.

The reader may wonder whether the functional representation given in the article is rich enough to explain the time shape of new product penetration: one might speculate about how additional variables, such as price and promotion, might be incorporated into the authors' simple model.

In "The Pattern of Consumer Purchases," A. S. C. Ehrenberg discusses two alternative models for explaining brand purchase behavior. The

first is based upon the mathematical function known as the "negative binomial distribution." If consumer purchasing behavior can be shown to follow a relatively simple mathematical law (for example, the negative binomial function) then the parameters of the function tell one all he needs to know about behavior in a new situation. These parameters can usually be estimated using far less data than would be required to trace out the entire distribution: the mathematical function summarizes the information obtained from past studies and makes it available for supplementing current data. The reader should note that the strategy is the same as in the Fourt and Woodlock article (where the time shape of market penetration summarized the relevant factors from studies on past new product introduction); Ehrenberg uses a more complex function which provides a richer set of implications. In particular, the validity of the negative binomial model implies that the standard error of the average quantity bought in a given period of time can be estimated from the proportion of nonbuyers and the mean. This is important because both statistics tend to be readily available from ordinary market research reports.

Second, Ehrenberg shows that the negative binomial distribution can be derived from a simple probability model, which can be used to explain consumer purchase data. Since our long-run hope is to be able to explain behavior rather than merely summarize and predict it, the existence of a reasonable structural model lends support to Ehrenberg's conclusions. The student should compare Ehrenberg's probability model with the ones presented in the two subsequent articles.

The articles by Frank and Kuehn focus upon the causal mechanism underlying the process of consumer brand choice. They are concerned with measuring the probability that a family will purchase a particular brand in a product class (for example, coffee or frozen orange juice), given its history of recent purchases; information about the purchase histories of individual families is available from consumer panel data. Both authors note that the probability of purchasing a given brand tends to increase as the number of purchases of that brand by the family in the recent past increases. Two different hypotheses for explaining this tendency are advanced.

In "Consumer Brand Choice—A Learning Process?" Kuehn argues that much of this increase in purchase probability is the result of a process of associative learning under conditions of reward: if a consumer purchases a brand and likes it, the probability of repurchasing that brand at a later date will be increased. Similarly, the probability of the brand's purchase on the next shopping trip will be reduced if a different brand has just been purchased. Frank's article, "Brand Choice as a Probability Process," suggests that much of the observed learning relationship can be attributed to the aggregation of families with different sets of stable

probabilities (that is, individual families may differ from one another, but each retains a constant purchase probability through time).

The student is left to draw his own conclusions about the relative merits of the two hypotheses, and to consider further research that might discriminate between them. In doing so, the following points should be kept in mind: (1) Is the learning model intuitively appealing, and does it have operationally significant implications for the practicing businessman (why is it worth looking into)? (2) What other mechanisms besides learning might cause a shift in a family's purchase probability, and what influence would such factors have upon the interpretation of Frank's and Kuehn's conclusions? (3) Does the difference between the classical and Bayesian approach to statistical testing affect the interpretation of the findings? More about both of the approaches will undoubtedly be appearing in future publications.



Early Prediction of Market Success for New Grocery Products*

.....

LOUIS A. FOURT and JOSEPH W. WOODLOCK†

MANY LEADING AMERICAN GROCERY MANUFACTURERS DERIVE FROM ONE-half to three-fourths of their sales from items that did not exist prior to World War II. Yet a survey of 200 large packaged goods manufacturers reveals that four out of five new products placed on the market after 1945 failed.¹ A reliable method for early selection of the most promising fraction of innovations would eliminate much of the loss now incurred on failures.

This article reports progress in early prediction of success or failure for new grocery store items. The data were obtained from National Consumer Panel records for national launchings, but the method is applicable to consumer panels in test markets.

The time required for prediction by this method depends upon the average interval between purchases for repeat customers. It can be only a few weeks for a new brand of margarine, but is nearly a year for cake mixes. This contrasts with older methods, using sales volume alone, in which the median time required for decision of new product success has been estimated at 19–24 months after completion of full-scale launching.² The new method has been applied to items in the following product classes:

Shortening cake mixes
Foam cake mixes
Frosting mixes
Specialty desserts
Cookie mixes
Margarine

Ready-to-eat cereals
Food drinks
Canned fruit
Scouring pads
Detergents
Pet foods

* Reprinted from the *Journal of Marketing*, national quarterly publication of the American Marketing Association, Vol. XXV, No. 2 (October, 1960), pp. 31–38.

† Abbott Laboratories and Market Research Corporation of America.

¹ New Product Introduction, U.S. Small Business Administration, Management Series No. 17 (Washington, D.C.: Government Printing Office, 1955), p. 63.

² Arthur C. Nielsen, Jr., "Consumer Product Acceptance Rates," in *Consumer Behavior*, edited by Lincoln H. Clark (New York, Harper & Bros., 1958).

These predictions depend upon the collection and efficient use of detailed information on the factors underlying sales. Sales volume for any item is the product of number of customers times frequency of purchase times size of purchase.

Observation of each of these components is required for our method of prediction, plus a further analysis: separation of initial from repeat purchases, and observation of the developing structure of repeat buying. All this information can be obtained from a consumer panel.

Consumer panel analysts often have examined the penetrations and first repeat buying ratios of new products.³ By penetration is meant the proportion of households that make an initial purchase of an item; by first repeat ratio is meant the fraction of initial buyers who make a second purchase.

Once the initial hurdle of attracting triers has been passed, this ratio is the most important single clue to the future success of an innovation. Reasonably high values for this ratio are a necessary (but not a self-sufficient) condition for success. In our experience repeat ratios below 0.15 (that is 15 out of every 100 triers) almost always spell failure; some very successful items convert as many as half of all triers into repeat purchasers. This facet of the observation technique makes possible identification of the very successful and very unsuccessful at a reasonably early stage.

The new elements in the method reported here are the following:

1. Experience with a large number of earlier new products is used to pre-determine the general functional form or shape of the cumulative penetration as a function of time. Observations of penetration for the particular new product are then used to determine its unique constants. Having estimated these constants, we can then extend the penetration as far into the future (for the same market) as desired.
2. The first repeat ratio is applied to this extended penetration curve to derive a cumulative first repeat purchase curve.
3. Subsequent repeat ratios as needed are similarly applied in turn. These are the ratio of third purchases to second, fourth to third, etc. The actual number of such ratios used depends on the frequency of purchase. The values used for these ratios are not merely those achieved to date, but are estimates for later periods, allowing each group of buyers an opportunity to make a repurchase.
4. The time intervals between purchases and the average size of transactions are observed, the latter separately for new and repeat buyers, and applied to obtain volume estimates. This separation is sometimes vital, for repeat customers may buy the large economy size or in multiple units, while triers may start cautiously.

³ Stanley Womer, "Some Applications of the Continuous Consumer Panel," *Journal of Marketing*, Vol. IX (October, 1944), pp. 132-36.

PENETRATION PREDICTION

As indicated above, extension of penetration from observed periods through prediction periods establishes the framework for volume prediction for grocery products.

There are products such as durable goods, novelties, and certain cosmetics, whose marketing depends solely or mostly upon one-time sales for the particular style or model. For such products, penetration is the entire story; with suitable modifications the procedures indicated here can be applied to such products.

Observation of numerous annual cumulative penetration curves shows that (1) successive increments in these curves decline, and that (2) the cumulative curve seems to approach a limiting penetration less than 100 percent of households—frequently far less.

A simple model with these properties states that the increments in penetration for equal time periods are proportional to the remaining distance to the limiting “ceiling” penetration. In other words, in each period the ceiling is approached by a constant fraction of the remaining distance.

Such a model is illustrated in Table 1 and in Figure 1, where the

TABLE 1
SIMPLE x, r PENETRATION MODEL
(Example: $x = 40\%$, $r = 0.3$)

Time Period	Increments in Penetration	
	Formula	Numerical Example
1.....	$r(x - 0) = rx$	$0.3(40) = 12$
2.....	$r(x - rx) = rx(1 - r)$	$0.3(40)(0.7) = 8.4$
3.....	$rx(1 - r)^2$	$0.3(40)(0.7)^2 = 5.9$
i	$rx(1 - r)^{i-1}$	$0.3(40)(0.7)^{i-1}$

ceiling, x , is 40 percent, and the constant of proportionality, r , is 0.3. In the first time period, the number of new buyers is $0.3(40 - 0) = 12$ percent. In the second time period, the number of new buyers is $0.3(40 - 12) = 8.4$ percent. Each increment is simply $1 - r$ times the preceding increment.

Ratios of successive increments in penetration, such as $8.4/12$ or $5.9/8.4$ in Table 1, are fast simple estimates of $1 - r$.⁴ These ratios can be

⁴Efficient estimates of $1 - r$ and x and a generalization of this model are the subject of a paper presented to the Stanford meeting of the Institute of Mathematical Statistics, August, 1960, by Professor Frank Anscombe of Princeton University. He points out that successive increments can be considered as drawings from a multinomial distribution and presents maximum likelihood estimates of $1 - r$ and x that are also sufficient statistics; that is, they utilize all the pertinent information in the observations.

averaged and applied to the last observed increment repeatedly, to extend penetration as far as desired. This model turns out to be somewhat too simple, but its basic properties remain usable.

To be more realistic, consider the fact that different buyers purchase a product class and its individual brands at widely differing rates. Experience shows that, when buyers are grouped by purchase rates into equal thirds, typically the heavy buying third accounts for 65 percent of the total volume, the middle third for 25 percent, and the light third for only

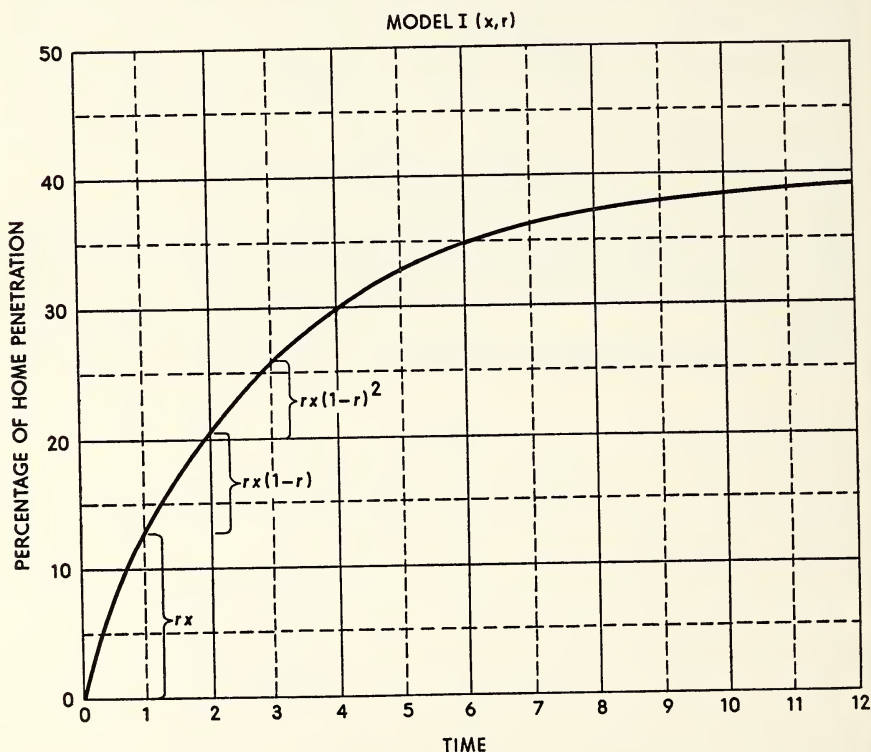


FIGURE 1. New buyer penetration—assumed.

10 percent. This means that, if transaction sizes are equal, heavy buyers make 6.5 purchases for every one of a light buyer, while medium buyers make 2.5, the total averaging 3.3.

If the original x, r model is applied to each of these thirds separately, this difference in purchase frequency will be sufficient to induce a remarkable "stretch-out" effect in the decline of increments of penetration for all buyers combined. This effect is sufficiently pronounced that the penetration model can be improved for the purpose of predicting a year ahead by assuming that increments of penetration approach a small positive constant, k , rather than zero.

A seemingly plausible alternative explanation for this behavior, panel

turnover, is ruled out by observations of penetration stretch-out in sub-samples chosen so that their composition does not change during the period of observation.

This second model is illustrated in Figure 2, where total penetration approaches a line whose value (at point t after i time periods) is $x_0 + ik$. We can convert our data to observations appropriate to the simpler two-parameter model by subtracting k from observed increments. The x in this simpler model then becomes the x_0 of the x, r, k model. Like x and r , k depends upon the individual item and its retail availability.

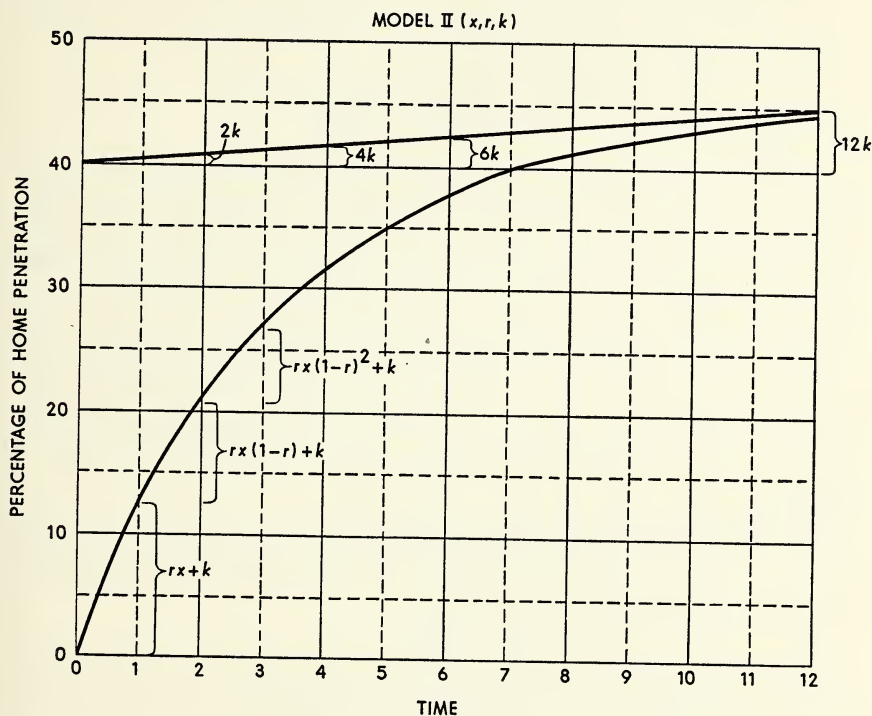


FIGURE 2. New buyer penetration—actual.

An empirical rule has worked rather well for estimating k : let k be one-half the increment of new buyers during the fourth average purchase cycle. The effect of subtracting too large or too small a k (when this is the largest source of error) is to produce characteristic serial correlation in the deviations of observations from the fitted curve of penetration. Too small an estimate for k causes the extreme (first and last) observed penetrations to exceed the fitted values, and those in the middle to be less than fitted values. Too large a k produces a reverse (excess) curvature.

Thus a means exists for detecting and correcting a mistake in the estimation of k . Typically k is a small number, of the order of 0.2 percent

per month, or 100,000 households for a nationally distributed grocery product. Actually k has not exceeded 200,000 in our experience, nor is the model very sensitive to small errors in k .

PREDICTIONS OF SALES

If sales volume in a second period is to be estimated from observations on sales in a first period, several assumptions are required. In market areas to be predicted:

1. Distribution will not shift greatly from the level existing at the end of the first period.
2. Promotional expenditures will not be substantially different in the second period from those during the latter part of the first period.
3. Prices will not change markedly.
4. Neither the product nor the package will be changed.
5. Competitive activity will not differ strikingly.

A shift in a single one of these observable factors should cause the prediction to be off in a predictable direction. A sixth assumption might be added—that the manufacturer does not know the prediction and hence does nothing to alter it. Altered behavior, of course, is a large part of the value of predictions, and failure for this reason is truly success.

Markets grow by deepening—through acquiring new triers and through developing repetitive buying in the original areas. Markets also grow by entry into new areas. Thus, the first assumption of constant distribution has the double burden of meaning steady availability in the old areas and of warning that new areas must be estimated as a separate procedure.

In an effort to predict as early as possible, we may want to use an observation period during which distribution is still increasing. This problem of gradual regional introduction can be minimized by observing various regions or markets separately.

The second assumption acknowledges that the costs of introductory promotions typically exceed levels that are profitable to maintain later. Hence, it is assumed that future promotion will be near current levels in the regions of earliest introduction—that is, levels current in those regions at the time of prediction.

Prediction of Repeat Ratios

The fraction of new buyers who have made a second purchase by the end of an observation period is necessarily an underestimate of those who will ever make a repeat purchase—for some have not yet had an opportunity to repeat.

This error can be substantially reduced by omitting the most recent new buyers from the denominator of the repeat buying ratio estimate. Omission of those purchasing within the most recent one or two average purchase cycles works well empirically. The first new buyers of an item

are typically heavy buyers of the product class. Their average purchase cycle is about one-half that of all buyers. (Compare 6.5 purchases with 3.3 for all buyers.) Thus, for equal observation times, the average purchase cycle of repeat buyers to date will be a smaller fraction of the eventual average if the product is purchased infrequently. Such products will require omission of two average purchase cycles' increments of new buyers while omission of one may suffice for frequently purchased product classes.

In practice, the decision to use one or two purchase cycles can be guided by the trends in the two repeat ratio estimates that result. Typically the trends converge as information accumulates.

Similar remarks hold for estimation of the proportion of second-time buyers who will ever make a third purchase, etc.

As might be expected, each successive purchase increases the probability of still another purchase. This is similar to the phenomenon noted by Alfred Kuehn for successive purchases of established brands; each consecutive purchase of a brand increases the probability that the next purchase by the household will be the same brand.⁵ Here the purchases do not have to be consecutive. Intervention of purchases of other brands merely prolongs the time until a given number of buyers make their n^{th} repeat purchase. While these interventions are important for some purposes, they need not be considered here.

Example of Application

Table 2 illustrates the process of obtaining estimated ratios for one product, and Table 3 their application. Together they present one case history in which data for a preliminary observation period were used to predict purchases and volume in a second period.

TABLE 2
DERIVATION OF REPEAT RATIOS

Buyer Type	Number of Purchases during Observation Period (000's)	Average Interval until Next Purchase (Months)	Number of Cycles in Lag	Number of Purchases in Observation Period Less Lag	Repeat Ratio
New buyers.....	6,021	2.41	2	4,472	
1st repeat.....	2,170	1.72	1	1,932	0.485
2d repeat.....	1,076	1.43	1	917	0.559
3d repeat.....	591	1.20	1	550	0.645
4th repeat.....	326	1.19	1	282	0.593
5th repeat.....	223	1.19	1	190	0.797
Over 5.....	627				3.300

⁵ Alfred Kuehn, "An Analysis of the Dynamics of Consumer Behavior and Its Implications for Marketing Management" (unpublished Ph.D. thesis, Graduate School of Industrial Administration, Carnegie Institute of Technology, 1958).

TABLE 3
ESTIMATED AND ACTUAL PURCHASES (r, x, k Model)

<i>Buyer Type</i>	<i>Estimated Number at End of Prediction Period (000's)</i>	<i>Number in Preliminary Observation Period (000's)</i>	<i>Difference or Estimated Addition in Prediction Period (000's)</i>	<i>Actual Addition (000's)</i>
New buyers.....	8,141	6,021	2,120	2,544
1st repeat.....	3,948	2,170	1,778	1,523
2d repeat.....	2,207	1,076	1,131	858
3d repeat.....	1,422	591	831	569
4th repeat.....	841	326	515	398
5th repeat.....	671	223	448	283
Over 5.....	2,214	627	1,587	1,711
Total purchases.....	19,444	11,034	8,410	7,886

The first column of Table 2 represents the observed number of purchases through the end of the observation period, in this case a year. The fourth-column purchases are observations through somewhat shorter periods. While 6,021 new purchases were made in the first year, only 4,472 had been made 4.82 (2×2.41) months before the end of the year. Only these are assumed to have had an adequate opportunity to make a second purchase. Similarly, although 2,170 first repeat purchases were made within the first year, only 1,932 had been made 1.72 months prior to the end of the period, and only these are regarded as having adequate opportunity to make a second repeat purchase.

The first repeat ratio, 0.485, is obtained by dividing 2,170, the number of first repeat purchases, by 4,472, the number of new purchasers who had an adequate opportunity to make a first repeat purchase. Second repeat and further ratios are obtained by using single cycle lags. Thus, 0.559 is the ratio of 1,076 to 1,932. Each new product has its own set of ratios. Comparison with ratios of other new products aids in the evaluation of a new item.

The final line in Table 2 represents a grouping of all purchases beyond the fifth repeat rather than the much smaller number of buyers. The value 3.3 is obtained by dividing 627 by 190. It is not mandatory that this ratio be used as the estimate of future purchases beyond the fifth repeat; work is in process to evaluate this ratio as a function of time.

The ratios in the final column of Table 2 are applied sequentially to the estimated number of new buyers reached by the end of the second period—in this example, 8,141. (This number comes from the application of the second penetration model— r, x, k .) Thus, $8,141 \times .485$ equals 3,948, the number of repeat buyers by the end of the second period. Similarly, $.559 \times 3,948$ yields 2,207, the number of second repeat buyers. It can be objected that these repeat ratios should be applied to lagged values of the previous waves of buyers. This is true, but in most instances the differ-

ences that result are minor because the cumulative buying curves are rather flat and parallel by the end of the second period.

The first column of Table 3 shows these estimates of purchases by type through the end of the second or prediction period. The second column, repeated from Table 2, shows actual purchases through the observation period. Because the method produces estimates on a cumulative basis through the end of the second period, it becomes necessary to subtract the actual results at the end of the first period to establish net additions during the second period.

The third column is the difference in the first two and represents predictions for the second period—again in this case a year.

The final column of Table 3 reports the actual number of buyers of each type in the second period as reported by the National Consumer Panel.

From the structure of Table 3, it can be seen that errors in prediction of one type of buyer may be partially compensated by offsetting errors for other types of buyers. In this example, under-estimates for new buyers and repeats beyond 5 help to compensate for over-estimates of the first 5 waves of repeats. In this sense the entire procedure is rather insensitive to error in individual details.

Multiplication of new buyers by their average transaction size, 1.05, and repeat buyers by theirs, 1.09, and adding, yields a total estimated package volume of 9,082,000 in the second period, or a predicted decline of 23 percent. Actual volume was 8,604,000—down 27 percent.

The example shown in Tables 2 and 3 is by no means the most accurate in our collection of case histories. It was chosen to present in detail because it illustrates difficulties as well as success in prediction.

Table 4 summarizes estimates that were made for this and six other

TABLE 4

COMPARISON OF PREDICTIONS WITH ACTUAL RESULTS FOR SEVEN PREPARED MIX ITEMS

Product	2d Period Volume Change		1st Period Repeat Volume as % of Total	2d Period Repeat Volume as % of Total	
	Estimate	Actual		Estimate	Actual
1.....	+ 5%	+ 1%	35%	64%	61%
2.....	-23	-27	47	77	69
3*	-40	-58	38	65	58
4.....	-51	-46	33	60	55
5.....	-53	-48	34	50	50
6†	-56	-48	46	76	74
7.....	-59	-59	24	47	‡

* Loss in distribution.

† Product reformulated.

‡ Not determined.

Items in Table 4 represent five different introduction times, four different price levels, and three different first-period volumes. In the two instances where actual change differed from predicted change by more than five percentage points, the stated assumptions were violated.

prepared mix items. Two marketing considerations deserve emphasis in considering such estimates:

1. Is the total volume predicted sufficient for profitable operation?
2. Is a sufficient part of that volume expected to come from repeat customers?

New buyer volume can be expected to deflate substantially through time. Repeat buyer volume is the ultimate determinant of success or failure.

These marketing considerations suggest that comparisons of predicted and actual results should be directed in the first instance to these questions:

1. Is the total volume estimate reasonably accurate?
2. Is the proportion of that volume from repeat customers predicted closely?

APPLICATION TO MARKET DEVELOPMENT

The foregoing method presents a reliable and easily usable prediction model for test markets or initial national marketings. It has the desirable feature of separating the very good and the very bad quickly. Intermediate cases can be observed for fairly long periods of time. In evaluating these longer observations, the model makes effective use of many previously neglected aspects of the cumulative marketing information available from panels.⁶

Finally, the model acts in a diagnostic capacity for failures. Too few triers might be due to the limitations of the promotional campaign or some limiting aspects of the labeling, naming, or use suggestions. Too small a repeat ratio implies that something is wrong with the product or the entire concept. Too long an interval between purchases demonstrates the need for large numbers of triers if product volume is to reach a reasonable level. This suggests the advisability of rapid introduction and greater stress of multiple uses—if any are available.

Of course, this work is too new for all possible difficulties to have been encountered yet. Regional variation in product acceptance is one problem; this can be handled by treating regions as separate markets. Development of a small but very loyal hard core of multi-repeat buyers is another problem. Tracing through many waves of repeat buying rather than grouping beyond some selected level permits treatment of this type of problem.

⁶ A different but complementary approach to new product evaluation using other aspects of panel information is contained in "The Dynamics of Brand Loyalty and Brand Switching," a paper by Dr. Benjamin Lipstein at the 5th Annual Conference of the Advertising Research Foundation, September 25, 1959, New York City. In particular, Dr. Lipstein's technique reveals the source of new buyers in terms of their earlier purchases.

The Pattern of Consumer Purchases*

.....

A. S. C. EHRENBURG†

GENERAL INTRODUCTION

Consumer Purchasing Data

WE ARE CONCERNED HERE WITH THE ORDINARY PRIVATE CONSUMER'S purchases of non-durable consumer goods. These goods are usually characterised by being marketed in pre-packed and branded form. Data about such purchases are obtained by market research techniques such as, in the case of Attwood's, the continuous consumer panels (see reference 1) based on random samples of either households or individuals and operated in various European countries including Great Britain. (The samples used in market research are almost always large in the statistical sense, so that no small-sample theory is required.)

The basic unit of time for measuring consumer purchases is usually a week, one week being generally like another. Most analyses are, however, made over periods of 4 or 13 weeks. For any such period of time, we therefore know how many consumers in the sample bought 0, 1, 2, 3, 4, or, in general, r , units of the given product, i.e. we know the frequency distribution of purchases. We generally also know what each of the consumers bought in preceding periods and can continue to watch his subsequent purchases.

The problem considered in this paper is the fit of the negative binomial distribution to such data. Product-fields analysed include the following:

Bread, Breakfast Cereals, Canned Vegetables, Cat and Dog Foods, Cocoa, Coffee, Confectionery, Detergents, Disinfectants, Edible Fats, Food Drinks, Household and Toilet Soaps, Jams and Marmalade, Polishes, Processed Cheese, Sausages, Shampoos, Soft Drinks, and Soups.

* Reprinted from *Applied Statistics*, Vol. VIII, No. 1 (March, 1959), pp. 26-41. Based on a paper read to the Study Section of the Royal Statistical Society on April 24, 1957.

† Research Services Limited, London.

The Negative Binomial Distribution

The Negative Binomial Distribution is a two-parameter distribution for the non-negative integers 0, 1, 2, 3, 4, and, in general, r . If the two parameters are taken as the mean m and the exponent k , the probability p_r of observing a number r is

$$p_r = \left(1 + \frac{m}{k}\right)^{-k} \frac{\Gamma(k+r-1)}{\Gamma(r)\Gamma(k-1)} \left(\frac{m}{m+k}\right)^r$$

These probabilities derive from the expansion of the binomial expression $[1 - m/(m+k)]^{-k}$ in which the exponent has a negative sign. (In contrast to the *positive* binomial distribution however, this mode of derivation does not seem to have any practical meaning.) It is often convenient to use instead of k the parameter $a = m/k$.

The negative binomial distribution is always positively skewed. It has one mode, which is at zero for the fairly small values of m and k which occur with consumer purchasing data, so that the distribution is then reversed-J-shaped. The variance of the distribution is

$$m(1 + m/k) = m(1 + a)$$

This formula and the higher moments can easily be derived from the characteristic function of the distribution. The general sampling theory of the distribution has been discussed by Anscombe,² who also gives earlier references.

Fitting the Negative Binomial Distribution

In the fitting of a negative binomial distribution to empirical data, the best estimate of the mean is the sample mean, since it is the maximum-likelihood estimator and is also unbiased, but the maximum-likelihood equations for a or k are very cumbersome to solve. Other ways of estimating a or k have however been developed (see reference 2), two of which are relevant here.

One way, the method of moments, is to estimate a by equating the observed sample variance to its expected value $m(1 + a)$. But this method of estimation is not particularly efficient. (The efficiency is sometimes less than 50%.) In any case, it would be laborious to have to compute the sample variance, especially since in market research the basic frequency distribution is often not tabulated.

The second and more attractive method is to equate the observed proportion of zero readings p_0 to its expected value, i.e. to write

$$p_0 = (1 + m/k)^{-k} = (1 + a)^{-m/a}$$

This equation can easily be solved by iteration, especially if it is written in the form suggested by Evans,³

$$a - c \cdot \ln(1 + a) = 0$$

where $c = -m/\ln p_0$. (Evans has also prepared a special table from which $\sqrt[3]{a}$ can be read off directly if it is entered with the appropriate value of $\log c$.) This method of estimation is very convenient for market research data, where the mean m and the proportion of non-buyers p_0 are often all the figures tabulated. It is at least 90% efficient for most of our kind of data, and often even a good deal more so.

Estimates of the probabilities p_r which may be required tend to be tedious to compute if the effective range of r is at all large. Probably the simplest procedure is to use the iterative formula

$$p_r = \left(\frac{a}{1+a} \right) \left(1 - \frac{a-m}{ar} \right) p_{r-1}$$

The goodness of fit of the estimated distribution can be tested by calculating the value of χ^2 for the observed and theoretical frequencies; but a quicker test is to compare the observed variance with the theoretical value $m(1+a)$, if the distribution has been fitted by the mean and the proportion of zeros. The variance of the differences between the observed and theoretical variances is (see references 2 and 3)

$$2m(m+a)(1+a) \left\{ \frac{(1+a)^2 \ln(1+a) - a(1+2a)}{(1+a) \ln(1+a) - a} \right\} \frac{1}{n} \\ + \frac{(1+a)a^4}{[(1+a) \ln(1+a) - a]^2} \{ (1+a)^{1+(m/a)} - (m+1+a) \} \frac{1}{n}$$

where n is the sample size. Evans³ has plotted this cumbersome expression for selected values of m and a in such a way that the standard error can be readily read off.

The Use of Transformations

Skew distributions are nowadays often transformed to simplify the subsequent analysis of the data, for descriptive purposes and from the point of view of sampling theory. Two closely related transformations are known for negative binomial distributions,⁴ namely

$$\sinh^{-1} \left(\frac{r+d}{k-2d} \right) \text{ and } \ln(r+k/2)$$

where in the first case d is an arbitrary constant which is optimum at $d = \frac{3}{8}$ if m is large and $k > 2$. Both transformations tend to stabilise the variance, at least for $k > 1$.

Transformations have, however, not so far been found useful in analysing consumer purchasing data. The reasons for this are threefold, namely that *arithmetic* means or totals are generally meaningful, that samples are usually large (so eliminating the more complicated statistical consequences of variance heterogeneity), and that the values of k are generally small (so that the known transformations do not in any case stabilise the variances very adequately).

The Fit of the Negative Binomial Distribution to Consumer Purchasing Data

In fitting the negative binomial distribution to consumer purchasing data in the product-fields mentioned above, a good fit has been obtained in most cases. A typical example, taken from the earliest data analysed, is shown in Table 1. More generally, Fig. 1 illustrates the goodness of fit for

TABLE 1
A TYPICAL EXAMPLE OF THE FIT OF A NEGATIVE BINOMIAL DISTRIBUTION
(26-Weekly Data for a 2,000-Household Sample)

Number of		Frequencies		Number of		Frequencies	
Units Bought	Observed	Theoretical	Units Bought	Observed	Theoretical	Observed	Theoretical
0.....	1612	1612	14.....	..	1.8	..	1.8
1.....	164	156.9	15.....	..	1.5	..	1.5
2.....	71	74.0	16.....	..	1.2	..	1.2
3.....	47	44.2	17.....	2	0.9	2	0.9
4.....	28	29.2	18.....	..	0.8	..	0.8
5.....	17	20.3	19.....	..	0.6	..	0.6
6.....	12	14.7	20.....	1	0.5	1	0.5
7.....	12	10.8	21.....	..	0.4	..	0.4
8.....	5	8.2	22.....	2	0.3	2	0.3
9.....	7	6.2	23.....	..	0.3	..	0.3
10.....	6	4.8	24.....	..	0.2	..	0.2
11.....	3	3.8	25.....	1	0.2	1	0.2
12.....	3	2.9	26.....	2	0.1	2	0.1
13.....	5	2.3	27 +....	..	0.9	..	0.9

$m = 0.636$, $p_0 = 0.806$, $k = 0.115$, $a = 5.53$
Standard Deviations: Root-mean-square 2.12, $\sqrt{[m(1+a)]} = 20.4$

a large number of cases, comparing the "theoretical" standard deviations $\sqrt{[m(1+a)]}$ with the "observed" root-mean-square deviation, a being calculated from p_0 and m .

The distribution fails, however, to give a good fit for distributions with relatively large standard deviations (about 5 or more), where the theoretical standard deviations tend to be somewhat too large. (See Fig. 1 and also Table 2.) This means that the observed distributions are not quite as skew as the negative binomial ones with the same values of m and p_0 . Distributions with large standard deviations tend to be those with large means, i.e. those for products which are very heavily bought, as for example margarine, detergents, and bread. It should be noted that the bias in the standard deviations is always positive and reasonably consistent. This suggests that it should be possible to find another class of distribution which will fit these data better.

An occasional feature of consumer purchasing data which is shown by the data in Table 1 may be pointed out here. It is the slight clustering at or just below the number of units equal to the number of weeks

covered (here 26), or at integral multiples of that number. This arises because some consumers will tend to buy practically the same number of units nearly every week. (Note also in Table 1 the suggestion of a cluster at 13.)

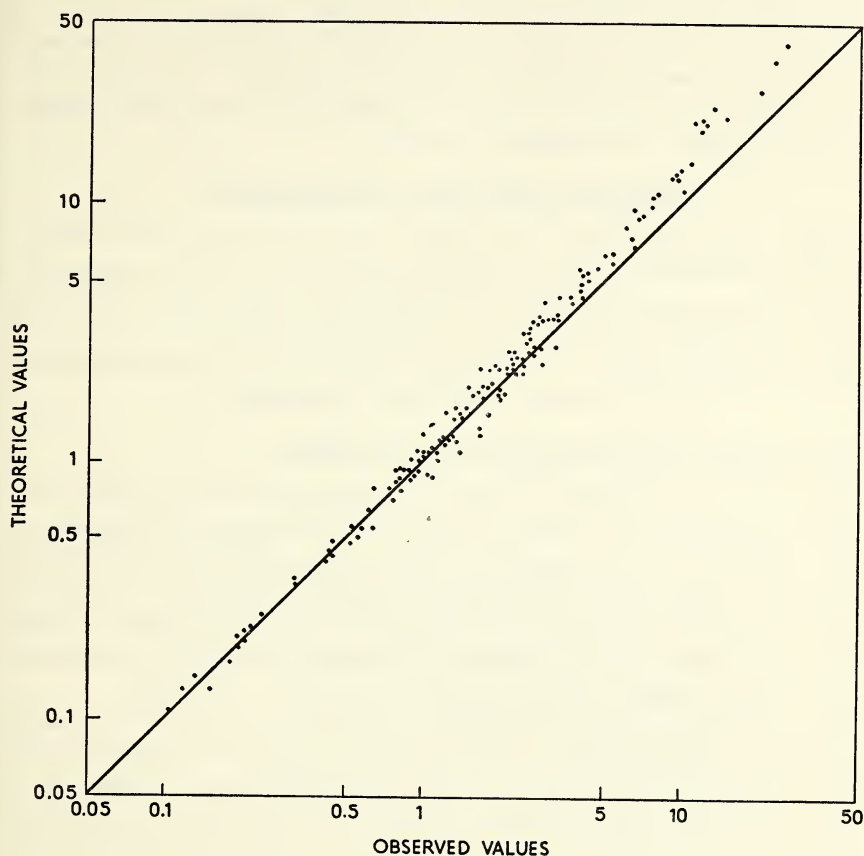


FIGURE 1. Comparison of "theoretical" and "observed" values for the standard deviation of the frequency distributions of consumer purchases.

Different Package Sizes

In fitting a discrete distribution like the negative binomial, the purchasing data must be expressed in terms of a basic unit. But many non-durable consumer goods are of course marketed in two or more package sizes, or even sold loose. So far it has been found that if such different package sizes are converted into equivalent units of a single "basic" size, a distribution can be adequately fitted and analyses such as estimating standard errors (see below) carried out. When there is a very popular size this may be used as the "basic" unit, or the larger size when there are two popular sizes, and so on. The choice of the basic unit is particularly important for some of the results which are given in the last section of this

paper. When there is any doubt about the appropriate basic unit to use, it is best to work out the results using each of the possible basic units. This will indicate the range within which the true value will lie, and it will usually also be obvious whereabouts in that range it must lie. So far, such ranges have usually been found to be fairly narrow. It must however be stressed that the occurrence of different package sizes can impose a definite, and sometimes an intractable, obstacle to the application of the negative binomial distribution to consumer purchasing data, except when each package size is treated separately.

IMMEDIATE USES AND CONSEQUENCES

Perhaps the first thing to say about the tendency for the negative binomial distribution to fit consumer purchasing data is that it is a pleasant and reassuring result. It is difficult to make this feeling explicit, but most people who handle extensive empirical data draw comfort from any evidence of systematic pattern which they come across. A number of more specific uses and consequences will now be discussed.

Quick Estimates of the Standard Error of the Mean

The fit of the distribution makes it possible to estimate quickly and easily the standard error of the average (or total) quantity bought in a given period of time. Instead of having to work out the sums of squares, etc., from a specially tabulated frequency distribution, the standard error can be estimated from the proportion of non-buyers p_0 and the mean m itself, the two statistics which tend to be regularly provided in the relevant market research reports, as already mentioned.

The methods mentioned earlier for estimating k or a can be used (e.g. Evans's special table), the standard error of m for a sample of n then, of course, being $\sqrt{[m(1+a)/n]}$. Alternatively, we find that it is quicker still and quite accurate enough for most purposes to use a nomogram from which the coefficient of variation of m can be read off directly for given m and p_0 . Calculating the sampling error of the mean m is therefore of an ease not far removed from that of using the well-known (positive) binomial formula for the error of the proportion p_0 of non-buyers.

Note that although for very heavily-bought products the negative binomial formula will over-estimate the standard error (see Fig. 1), this does not necessarily matter greatly. In market research, an indication of the *order* of the sampling error is often all that is required, and some allowance for the bias can in any case be made from the general trend illustrated in Fig. 1.

Outlying Values

The fit of the negative binomial distribution makes it possible to investigate the statistical status of outlying values, in particular perhaps

that of the occasional very large purchase. (Indeed, it was in such a connection that the work which has led to this paper was first undertaken, following a suggestion by Mr. D. A. Brown.) This type of problem can probably be best explained in terms of an example.

In the Attwood Consumer Panel in Western Germany it was found that one or two households in any one period would buy what seemed to be abnormally large quantities of a certain class of detergents. When negative binomial distributions were fitted to the data, the theoretical standard deviations were significantly smaller than those actually observed, thus showing that these large purchases were in fact larger than the distribution of all the other purchases would appear to warrant, on the basis of the negative binomial distribution. This obviously does not show that these very large purchases were in any sense "wrong," but it does show that they are statistically abnormal, since detergent purchases in other countries, and purchases in other fields in Western Germany, did not show this deviation from the negative binomial form. Such a conclusion does not of course tell us what to do about the data, but it is at least useful in such cases really to know that the readings *are* statistically abnormal, instead of merely having a subjective opinion to that effect.

Further analysis showed that the major individual brands of detergents did not show such abnormalities at all, and that we were dealing with very heavy purchases of some relatively minor brand or brands. When the back-ground of the purchases was investigated, several "abnormal" points were found. One, for example, was that certain of the lesser-known and relatively cheap brands are sold by door-to-door salesmen, and that these brands are then occasionally bought in very large quantities. Clearly, we are dealing with two types of purchase, namely those made from the ordinary retail outlets and those made at the door. It was therefore reasonable that the negative binomial distribution analysis should show an abnormality in the pattern of purchases before these two types of purchases were differentiated.

The Fit in Sub-Groups of the Population

In market research one is usually interested not only in a given population of consumers as a whole, but also in sub-groups, such as consumers in specific age-groups or in different geographical regions. The question then arises whether the negative binomial distribution will give a good fit for such sub-groups when it does so for the population as a whole, or vice versa.

In theory, different negative binomial distributions, variously weighted and added together, will practically never combine into an exact negative binomial distribution. But in practice, negative binomial distributions for sub-groups do seem to combine into a negative binomial distribution for the population as a whole, the fit in all cases being approximate rather than exact. This is illustrated in Table 2, where for the lightly-bought

TABLE 2

NEGATIVE BINOMIAL PARAMETERS IN DIFFERENT SIZE-OF-HOUSEHOLD GROUPS,
FOR PURCHASES OF A LIGHTLY BOUGHT AND A HEAVILY BOUGHT BRAND

Brand	Parameters	Size of Household Groups					Total Sample
		1	2	3	4	5 +	
Light	m	0.19	0.42	0.55	0.60	0.69	0.51
	$10k$	0.32	0.42	0.63	0.76	0.71	0.58
	a	5.9	10.1	8.6	7.9	9.8	8.8
	Standard Deviations:						
	Root-mean-square	1.2	2.2	2.5	2.3	2.9	2.4
	$\sqrt{[m(1+a)]}$	1.2	2.2	2.3	2.3	2.7	2.2
Heavy	m	2.8	6.3	8.5	12.6	18.2	16.0
	k	0.24	0.21	0.18	0.21	0.18	0.19
	a	12	30	46	59	102	50
	Standard Deviations:						
	Root-mean-square	4.6	9.6	12.9	20.7	26.2	16.0
	$\sqrt{[m(1+a)]}$	5.9	14.0	19.9	27.7	43.2	22.0

brand the distributions of the purchases made by households of sizes 1, 2, 3, 4, and 5+ respectively, and by all households together, are all negative binomial, as is shown by the agreement of the two estimates of the standard deviations. Even for the very heavily-bought brand in Table 2, where the fit of the theoretical distribution is not so good, it is clear that the observed distributions in the sub-groups and in the population as a whole are of the same kind.

The explanation of this paradox lies mainly in the fact that even for something as effective as a breakdown by size-of-household, the differences between the distributions for the sub-groups are small compared with the scatter within each distribution. Furthermore, the breakdown by size-of-household is not untypical of others in market research in that much the largest sub-groups are those of sizes 2, 3, and 4, which are in any case relatively similar to each other. (In other words, combining sizes 1 and 5+ only, for example, might well not lead to a good negative binomial fit; but this would not be of practical interest.)

The Fit for Individual Brands and Product-Groups

A superficially similar problem, though technically diametrically opposed, arises when we consider the fit of the distribution for individual brands on the one hand and for combinations of brands (e.g. the whole product-group) on the other.

Here we have first of all a theoretical result which can easily be derived from the characteristic functions, namely that if x and y are two negative binomial variables with parameters (m_x, k_x) and (m_y, k_y) , then their sum $(x + y)$ will follow a negative binomial distribution

with parameters $(m_x + m_y)$, $(k_x + k_y)$, if $a_x = a_y$ where $a = m/k$ as usual.

This is of course a rather limited result even though one may expect it to hold approximately when a_x and a_y are not exactly equal. But from the practical point of view, one perhaps need not expect much trouble when looking at the distribution of the purchases of two (or more) brands. For either the two brands are more or less equally heavily bought so that their distributions (including the values of a) probably do not differ very much from each other, or one brand is much the less popular, in which case it will in any event not markedly influence the distribution of the two brands combined.

Mainly for practical reasons not much numerical work has yet been done on this topic. But the whole question of comparing the purchasing patterns of different brands in the same product field which is here opened up is obviously one of great interest.

Constant Parameters

In the theoretical results just mentioned, the parameter a had to be the same for two different negative binomial variables if the sum of the variables was to follow a negative binomial distribution also. It is therefore interesting that in such limited work as has so far been done, the parameter a has also been found to be fairly constant under two kinds of empirical conditions for a given brand, namely:

1. When there is movement (e.g. seasonal) in the rate of buying, i.e. when the mean m varies from one period to the next.
2. When different sub-groups of the population have different mean rates of buying m . (See Table 2, the "Light" brand.)

Although the parameter a seems to possess little in the way of an immediately obvious physical meaning, this tentative finding that it might behave as an empirical constant for a given brand is highly promising and will have to be studied with some care. Note that if a is invariant, the parameter k must of course vary with m .

One definite exception to this tentative finding is so far known. It is that the invariance of a for different sub-groups does not hold for the very heavily-bought brands for which the negative binomial distribution itself does not, as already mentioned, give a good fit. An example is given in Table 2 (the "Heavy" brand), where k is fairly stable whilst a varies with m . This difference may perhaps offer a line of attack for studying the poor fit of the negative binomial distribution for such heavily-bought brands.

A SIMPLE STOCHASTIC MODEL

The Model

The negative binomial is a particularly interesting distribution because several theoretical models are known for its derivation (see reference 2). One of these models seemed relevant for consumer purchasing data and

this led us to study the fit of negative binomial distributions for such data. The model is a two-dimensional one, one dimension being time and the other (an unordered one) being the individual consumers, as follows (see also Table 3):

TABLE 3
THE STOCHASTIC MODEL OF THE NEGATIVE BINOMIAL
DISTRIBUTION FOR CONSUMER PURCHASES

Consumers	Periods of Time						Long-run Averages	Distributions (Horizontally)
	I	II	III	IV	.	.		
A	x	x	x	x	x	x	μ_A	Poisson
B	x	x	x	x	x	x	μ_B	Poisson
C	x	x	x	x	x	x	μ_C	Poisson
D	x	x	x	x	x	x	μ_D	Poisson
.	x	x	x	x	x	x	$\mu.$	Poisson
.	x	x	x	x	x	x	$\mu.$	Poisson
Mean	m	m	m	m	m	m	m	
Distributions (vertically)	Neg. Bin.	Neg. Bin.	Neg. Bin.	Neg. Bin.	Neg. Bin.	Neg. Bin.	χ^2	

Note: The x-values are given without subscripts; they represent the various entries in the table and are not intended to imply equality.

Poisson Distribution in Time. For the purchases of any particular consumer in successive periods of time, e.g. purchases of 2, 0, 3, 1, 1 units and so on, the model requires that these purchases behave like independent random samples from a Poisson distribution. This is plausible under two conditions which should normally be more or less fulfilled, namely

a) That the successive periods of time are not only of equal length but also similar to each other, e.g. weeks or longer periods measured in weeks, rather than days.

b) That the periods are not too short, so that the purchases made in one period do not directly affect those made in the next. Periods of one week for instance may be too short for some products: e.g. if a tin of cocoa is bought in one week, no such purchase is likely to be made in the following week.

The Poisson distribution has one parameter, the mean μ , i.e. the average rate of purchasing 'in the long run.'

A χ^2 -Distribution of Consumers. The second part of the model then specifies that the distribution of the average rates of purchasing μ of different consumers should be proportional to a χ^2 or Type III distribution with $2k$ degrees of freedom. This is also plausible, since such a distribution is fairly flexible (having two adjustable parameters) and of the right shape (i.e. a continuous distribution for non-negative values, reversed-J-shaped or hump-backed and always positively skewed).

Note that it is not necessary to assume that purchases actually follow this model in successive periods, let alone in the long run. Indeed, the average rates of purchasing μ which have been postulated for individual consumers need not be observable in any sense. It is only necessary to

suppose that in any *one* period of time purchases behave as if they were a random sample from such a model. In the next period the parameters of the model may well have changed, or the model might even have broken down altogether.

A number of deductions can be made from this model and tested empirically. If validated, such deductions will be of interest not only in their own right, but also because they then lend support to the validity of the model generally. (The first deduction of this kind which has been validated is of course that the negative binomial distribution does tend to fit consumer purchasing data, as we have seen.)

This model is of particular interest where continuous research information is available, i.e. the purchasing data of the same sample of consumers in different periods of time, as in the case of consumer panels.

The Standard Error of the Difference between the Mean Purchasing Rates in Two Periods of Time, if There Is No Trend

The trends shown in the course of time are one of the most important aspects of consumer purchasing data. In particular, we often have to compare the mean purchasing rate m' in one period of time with the mean purchasing rate m'' in another, equal, period. The standard error of the difference ($m' - m''$) can of course be computed from the individual readings, but the question arises whether it cannot be estimated more quickly and easily.

If the means m' and m'' are based on *independent* samples of size n' and n'' , it follows from the earlier results that a simple estimate of the standard error of ($m' - m''$) can be calculated from the means and the proportions of non-buyers in each of the two periods, in the form $\sqrt{[m'(1 + a')/n' + m''(1 + a'')/n'']}$. But if the same sample of consumers is used in both periods, as in work based on consumer panels, the two sets of readings will be correlated. To obtain an estimate of the standard error in that case, we shall suppose for the moment that the mean purchasing rates in the population are the same, i.e. that there has been no trend. Then if r' and r'' are the purchases of any one customer in the two periods, we require the variance of $(r' - r'')$ over all consumers in the population. This is equal to the expected value of $(r' - r'')^2$, since the mean of $(r' - r'')$ for all consumers has been assumed to be zero. Now in the model, r' and r'' for any one consumer behave like independent readings from the same Poisson law, so that his expected value of $(r' - r'')^2$ is equal to the expected value of $(r' - \mu)^2 + (r'' - \mu)^2$, which is twice the mean of the Poisson. It follows that for all consumers the variance of $(r' - r'')$ is equal to twice the average rate of purchasing in the population.

The sample estimate of this variance is $(m' + m'')$. The standard error of the difference between the means m' and m'' obtained from the same sample is therefore:

$\sqrt{[(m' + m'')/n]}$, or $\sqrt{(2m/n)}$ for short.

The values given by this theoretical formula compare well with the values of the standard error computed directly from the individual differences $(r' - r'')$, as shown in Fig. 2. It may be noted that the derivation of the formula $2m/n$ depends only on the Poisson part of the theo-

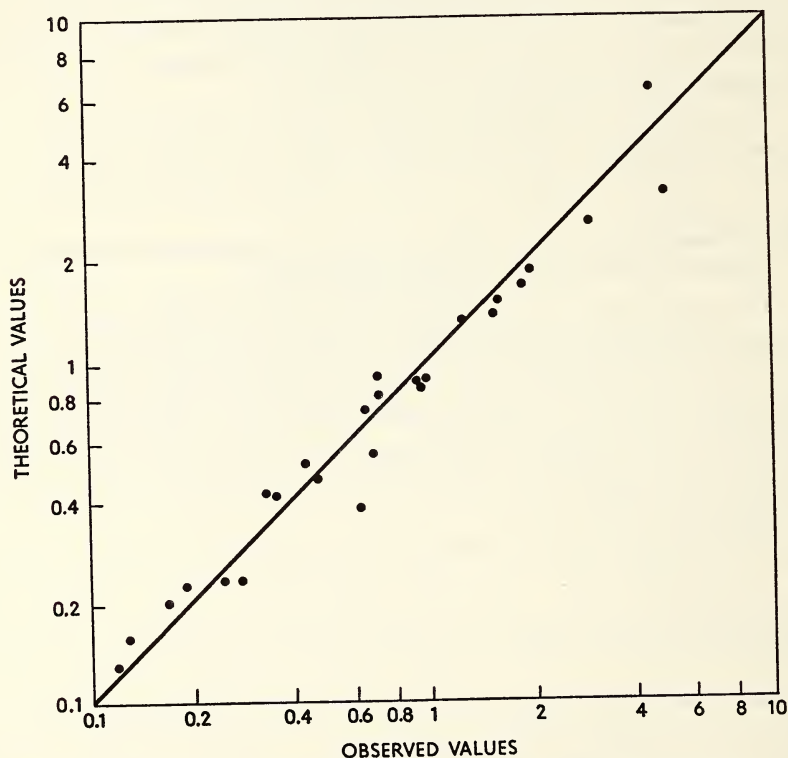


FIGURE 2. Comparison of "theoretical" and "observed" values for the standard error of the difference between estimates of mean purchasing rate in two periods of time.

retical model, and not on the χ^2 part. The fact that the formula holds even where the negative binomial distribution as a whole does not fit so well, e.g. for heavily-bought products, is therefore not surprising.

The Standard Error of the Difference between the Mean Purchasing Rates in Two Periods of Time, if There Is a Trend

In addition to knowing the standard error of the difference $(m' - m'')$, based on the same sample, when there is no trend, it is also important to know this standard error when m' and m'' are significantly different. For example, we may wish to know the limits of sampling error for an observed difference, or we may wish to compare the observed trend in purchases with sales or production statistics, or with the cor-

responding trend the year before, or with the trend in some other group of consumers.

The derivation of a theoretical formula for the standard error of $(m' - m'')$ when there is a trend is difficult because it will depend on the nature of the trend. A trend may occur because of a change in the proportion of buyers, or because of a change in the pattern of buying of the original buyers, or because of some combination of the two.

However, a starting point is indicated by the empirical finding already mentioned that the parameter a tends to be fairly stable even when there is a trend. We note that the correlation R between the purchases made in one period and those made in the other is $(a/1 + a)$ in terms of the negative binomial parameters, if there is no trend. (R is used as a symbol for correlation here because r has already been used for another purpose.) It follows then that if a is constant, the theoretical correlation formula appropriate to one period is equal to that appropriate to the other period. From this it is an easy step to suppose that the actual correlation between two periods would also be $(a/1 + a)$, despite the trend.

From this assumption it follows that the standard error of $(m' - m'')$ is

$$\sqrt{\{m' + m'' + a(\sqrt{m'} - \sqrt{m''})^2\}/n}$$

The interesting point about this formula is that the contribution of the term $a(\sqrt{m'} - \sqrt{m''})^2$ by which the formula differs from the earlier result $\sqrt{(2m/n)}$ will generally be very small or even negligible. For example, even when m' is as much as 50% higher than m'' , and a is as large as about 10 (which is fairly typical of the *larger* values of a), the value given by the above formula is only about 10% higher than that given by the formula $\sqrt{(2m/n)}$. This suggests that the assumption on which the above formula rests is not very important, and that generally even the simple formula $\sqrt{(2m/n)}$ should apply, to a good degree of approximation, for any reasonable kind of trend.

An approximate formula can also be derived for the special case where we compare one seasonal trend with another, the data being based on the same sample throughout. For if, in calculating the standard error formula, we can again ignore the presence of trends, the standard error of the difference between $(m' - m'')$ in one year and $(m' - m'')$ in the other year is then approximately $\sqrt{(4m/n)}$.

The Standard Error of the Difference between the Proportions of Non-Buyers in Two Periods of Time

The proportion of non-buyers p_0 (or its complement, the proportion of buyers) is an important quantity in market research. Its sampling error is given by the well known (positive) binomial variance formula, $P_0(1 - p_0)/n$. The variance of the difference $(p'_{0,i} - p''_{0,i})$ in the esti-

mated proportions of non-buyers $p'_{0,i}$ and $p''_{0,i}$ for the same sample in two different periods of time of length i is

$$(p'_{0,i} + p''_{0,i} - 2p_{0,2i})$$

or

$$2(p_{0,i} - p_{0,2i})$$

if there is no trend, where $p_{0,i}$ is the proportion of non-buyers in each period i and $p_{0,2i}$ is the proportion of non-buyers in both periods combined. This formula can also be put into the form

$$2(\text{Proportion of New Buyers} + \text{Proportion of Lapsed Buyers}),$$

where the New Buyers and the Lapsed Buyers are respectively the consumers who bought in the second period only and in the first period only. (These proportions are often known from so-called "Loyalty Analyses.")

These are general formulae. But in practice, information about $p_{0,2i}$ or about the new and the lapsed buyers is often not readily available. The variance of $(p'_{0,i} - p''_{0,i})$ can then be estimated from the negative binomial model, assuming no trend. In terms of the parameters a and k for a period of length i , the variance of $(p'_{0,i} - p''_{0,i})$ is

$$2[(1+a)^{-k} - (1+2a)^{-k}]$$

To calculate a and k takes time, whilst in certain types of market research investigation the information on the mean quantity m (required for estimating a and k in the way mentioned earlier) either may not be available or may not be very reliable. But if the proportions of non-buyers in two periods of time of different lengths, say 4 weeks and 13 weeks, are known, as is often the case, the above formula can be evaluated.

An explicit though approximate solution for the variance of $(p'_{0,4} - p''_{0,4})$, where the proportions of non-buyers $p_{0,4}$ for 4 weeks and $p_{0,13}$ for 13 weeks are known, is

$$2[p_{0,4} - \exp\{0.588 \ln p_{0,13} + 0.412 \ln p_{0,4}\}]$$

assuming a is reasonably large compared with unity. Further approximation reduces this formula to

$$2[p_{0,4} - \sqrt{(p_{0,4}p_{0,13})}]$$

This is an underestimate of the true variance. For the kind of purchasing data that occur in practice, the formula

$$2[p_{0,4} - 0.95\sqrt{(p_{0,4}p_{0,13})}]$$

is generally an overestimate, so that we have upper and lower bounds. For large values of $p_{0,4}$, a closer upper bound is given by replacing 0.95 in the above formula by 0.98 (when $p_{0,4}$ is about 0.8) or 0.99 (when

$p_{0.4}$ is about 0.9). Note also that for large $p_{0.13}$ or small $p_{0.4}$, the arithmetic mean of $p_{0.4}$ and $p_{0.13}$ can be used instead of the geometric one. Similar approximate formulae for quick application can be worked out for other periods of time.

The Effect of Partial Changes in the Sample

It will often happen that even though the sample used in two different periods is very largely the same, it will not be *exactly* the same. Some of the consumers used in one period may have been replaced in the other. (We must suppose that this replacement has been competently carried out, e.g. by random sampling subject to various stratifications.) There will then be an increase in the sampling error of the difference ($m' - m''$) of the mean purchasing rates, which can be readily allowed for. Thus, if the proportion of the sample which is *not* in common to the two periods is c , and if R is the correlation between the purchases of the same consumers in the two periods, then the standard error formula $\sqrt{(2m/n)}$ has to be multiplied by the factor $\sqrt{[(1 - R + Rc)/(1 - R)]}$, or $\sqrt{(1 + ac)}$ in terms of the parameter a .

The Length of the Period of Time Covered

A number of simple results concerning the length of the period of time over which purchases have been recorded also follow from the theoretical model. Consider two periods of time, one of unit length and the other L times as long, and distinguish the relevant parameters by the suffixes I and L .

If we assume no trend, it is obvious that $m_L = Lm_I$. But the parameter k will be the same in either period. Therefore $a_L = La_I$. The correlation R_L between two periods of length L will then be $R_L = LR_I / (1 - R_I + LR_I)$, in terms of the correlation R_I . The standard errors of m_I and m_L will be in the proportion of $\sqrt{[(1 + a_I)/L(1 + La_I)]}$, whilst the standard errors of the differences between means of two periods will be in the proportion $\sqrt{(1/L)}$.

Results of this sort can be useful for computational purposes and, more important, for the planning of new projects. For example, in planning research for a test marketing operation it might be necessary to know whether it is more accurate to ascertain the purchases of a relatively large number of consumers over a short period of time or those of a smaller number of consumers over a longer period.

The Need for a Non-Stationary Model

In theory, the various formulae which we have discussed for the analysis of purchases in two different periods apply equally well to periods which are some time apart as they do to successive periods, because the model of heterogeneous Poisson sampling which we are using is stationary in time.

But common sense suggests that increasing the distance between two periods will generally decrease the correlation. Thus, although the purchases of a certain brand by a given household may well follow a Poisson law for some time, the household may suddenly transfer its loyalty to a different brand altogether. The correlation for periods relatively far apart will then of course be reduced.

A subject for further development is therefore a model which, though in the first instance still stationary as regards the average purchasing rate, is not so as regards *individual* rates. Alternatively, such a model may be regarded as one in which successive readings for one consumer are no longer independent.

Some Mathematical Calculations

In conclusion, one or two consequences of the negative binomial distribution and its model will be mentioned which may be of a certain interest.

Suppose for example that for a given brand or product we know:

1. The proportion of non-buyers in a certain period of time.
2. The proportion of non-buyers in another, longer, period of time.

Then assuming that there has been no trend, for which one may have indirect evidence (e.g. from the proportions of non-buyers), we can obtain the average rate of buying and the parameters of the negative binomial distribution, i.e. we can calculate how much the people who did buy bought.

Again, suppose that for a given period of time we know only the proportion of non-buyers and the proportion of buyers of a single unit, we can calculate the proportions of buyers of two units, three units, and so on, and in particular, the average rate of buying.

Such results, and there are others, have a twofold interest, one possibly practical, the other methodological.

The practical implications are that the first result, for instance, could be used with the kind of market research surveys where it is asked: "Have you bought brand X in the last month, and if so, have you bought it during the last week?" The information so obtained can then be translated into estimates of the average amount bought. Similarly, whereas asking people in a one-shot survey how much of a product they have bought in a given period of time is generally unreliable, it may be less so if we simply ask whether, if they bought at all, they bought only one unit, from which the average rate of buying can again be calculated. This sort of analysis could also be used for experimental work on the memory factor.

The methodological implication is that we should be taking another step in the direction of not necessarily having to observe directly everything we wish to know, through being able to make theoretical calcula-

tions instead. So far, the biggest step of this kind in market research work has been the application of sampling theory. But it is to be hoped that with increasing knowledge and improved techniques, other more general kinds of applied mathematics may also become increasingly applicable.

REFERENCES

1. WADSWORTH, R. D. "The Experience of a User of a Consumer Panel," *Applied Statistics*, Vol. I (1952), p. 169.
2. ANSCOMBE, F. J. "Sampling Theory of the Negative Binomial and Logarithmic Distribution," *Biometrika*, Vol. XXXVII (1950), p. 308.
3. EVANS, D. A. "Experimental Evidence Concerning Contagious Distributions in Ecology," *Biometrika*, Vol. XL (1953), p. 186.
4. ANSCOMBE, F. J. "The Transformation of Poisson, Binomial and Negative Binomial Data," *Biometrika*, Vol. XXXV (1948), p. 246.

Brand Choice as a Probability Process*

RONALD E. FRANK†

INTRODUCTION‡

IT IS THE OBJECTIVE OF THIS ARTICLE TO ANALYZE THE CHRONOLOGICAL patterns of consumer brand choice of frequently purchased products such as food and household items. The ultimate goals of such an investigation are (1) to gain insight into the behavioral processes that underlie the observed patterns of brand choice; and (2) to provide a framework for predicting the effect on brand choice of such elements in the marketing mix as changes in relative price, distribution, and promotion.

This investigation aims to gain insight into the underlying processes that generate the observed patterns of brand choice for individual families by comparing their actual choice patterns to those generated by a simple chance model. This comparison is useful in evaluating at least one aspect of brand loyalty.

Despite the importance of predicting and influencing brand choice, only two writers have published material on the subject. These analyses, by George Brown¹ and Ross M. Cunningham,² have stimulated increasing interest in this area.

* Reprinted from the *Journal of Business*, Vol. XXXV, No. 1 (January, 1962), pp. 43-56. Copyright 1962 by the University of Chicago, all rights reserved.

† Harvard University.

‡ The study resulting in this publication was partly financed by fellowships granted by the Earhart Foundation and the Graduate School of Business of the University of Chicago. Harry V. Roberts, James Coleman, and Lester G. Telser provided continual counsel and encouragement. I am also indebted to William Kruskal for several fruitful suggestions. The conclusions and interpretations are those of the author and do not necessarily reflect the view of the above organizations and individuals.

¹ "Brand Loyalty—Fact or Fiction?" *Advertising Age*, Vol. XXIII (June 9, 1952), pp. 53-55; (June 30, 1952), pp. 45-47; (July 14, 1952), pp. 54-56; (July 28, 1952), pp. 46-48; (August 11, 1952), pp. 56-58; (September 1, 1952), pp. 44-48; (September 22, 1952), pp. 80-82; (October 6, 1952), pp. 82-86; (December 1, 1952), pp. 76-79; Vol. XXIV (January 26, 1953), pp. 75-76.

² "Brand Loyalty—What, Where, How Much?" *Harvard Business Review*, Vol. XXXIV, No. 1 (January-February, 1956), pp. 116-28; "Measurement of Brand

Recently, in 1958, Alfred A. Kuehn completed a dissertation³ which represents the first attempt to use probability models as a tool for analyzing consumer brand choice. Although the present study began independently of Kuehn's work, the framework used in analysis is quite similar to the one he used.

In what follows, there is first a description of the data upon which the analysis is based. Next, I discuss the conceptual framework used in the exploration. Finally, I give the operational definitions, the results, and the conclusions.

THE DATA

The analysis uses records of regular- and instant-coffee purchases of 536 continuously reporting families belonging to the *Chicago Tribune's* consumer panel from August, 1957, to September, 1958.

The *Tribune* panel consists of a number of families who keep a chronological record of their purchases of food and household items. For each purchase in a given product class, information is available as to the family code number, selected demographic characteristics of the family, brand purchased, date, quantity, price, type of outlet, and whether or not a deal⁴ was used in making the purchase.

More than twelve thousand purchases were made by the 536 families during the period studied.

Panel data provide a continuous record of brand choices for an extended period of time, which is essential for a study of short-run brand purchase patterns. The data were purchased from an existing panel because they were available at a low cost. The *Tribune's* panel was selected as the source because it covered a limited geographic area, namely the city of Chicago and principal suburbs within a forty-mile radius. Although one may not be justified in reaching sweeping generalizations from such data, there are advantages in a study of limited scope because fewer variables need be included.

FRAMEWORK FOR ANALYSIS

The following questions were investigated:

1. Given that a customer has made a run of 1, 2, 3 . . . n purchases of a certain brand, what is the relative frequency of buying the same brand at next purchase?

Loyalty," *The Marketing Revolution* (New York: American Marketing Association, December 27-29, 1955), pp. 39-45; and "Brand Loyalty and Store Loyalty Interrelationships," *Marketing Keys to Profits in the 1960's* (New York: American Marketing Association, June, 1959).

³ "An Analysis of the Dynamics of Consumer Behavior and Its Implications for Marketing Management" (unpublished Ph.D. dissertation, Graduate School of Industrial Administration, Carnegie Institute of Technology, May, 1958).

⁴ As used in the panel, a deal is any special sale at point of purchase (for example, a one-cent sale or a coupon).

2. Given that a customer left a certain brand 1, 2, 3 . . . n purchases ago, what is the relative frequency of buying it at the next purchase?

In the analysis that follows the relative frequencies mentioned in questions 1 and 2 are referred to as a brand's repeat purchase probability and return purchase probability, respectively.

This relationship is looked upon in two ways: first, as a measure of the extent to which a customer's brand choices in the immediate past can be used to predict the next choice; second, as a measure of the extent of habitual purchasing behavior. Habitual purchasing behavior can be thought of as one aspect of brand loyalty. For the purpose of this research, "habitual purchasing behavior" is defined as the existence of a positive association between the number of previous purchases of a particular brand and the consumer repeat purchase probability. For example, suppose Table 1 presents a hypothetical relationship between the repeat

TABLE 1
HYPOTHETICAL RELATIONSHIP BETWEEN PROBABILITY OF
BUYING BRAND A ON NEXT PURCHASE AND NUMBER
OF PREVIOUS CONSECUTIVE PURCHASES OF BRAND A

Cus- tomer	Number of Previous Consecutive Purchases of Brand A						
	1	2	3	4	5	6	7
1.....	0.30	0.30	0.30	0.30	0.30	0.30	0.30
2.....	0.30	0.46	0.59	0.65	0.73	0.79	0.81

purchase probability for Brand A and the number of previous consecutive purchases of Brand A for each of two customers. Customer No. 1 exhibits no association between the probability of buying Brand A and the number of previous consecutive purchases of the brand. That is to say, the probability of the customer buying the brand is independent of the number of past consecutive purchases. Customer No. 2, on the other hand, shows dependence in the sense that the probability of buying the brand increases as the number of previous consecutive purchases of the brand increases.

Customer No. 1 exhibits no habitual behavior, while Customer No. 2 exhibits a degree of habitual behavior.⁵

It is my purpose to analyze the panel data in a way that illuminates the nature of the processes that generate the observed patterns of brand choice. For example, the process generating the observed purchase pattern for Customer No. 1 exhibiting no habitual behavior might be characterized by the following urn model. An urn contains b black and r red balls. Drawing out a black ball represents a purchase of Brand b , while

⁵ In using his or any other definition of habit, it is useful to think of customers placed in a continuum, rather than being separated into qualitatively distinct classes.

drawing out a red ball stands for Brand r . A ball is drawn at random (i.e., a purchase is made at random). The ball is replaced before the next drawing. This process is repeated each time a purchase is made. The probability that Brand b will be purchased is always $b/b + r$. This probability is independent of the previous brand choice of the customer. This model corresponds with Customer No. 1 in Table 1.

In contrast, another urn model might be used as a basis for characterizing a habitual pattern of behavior. An urn again contains b black and r red balls. A black ball represents a purchase of Brand b , while a red ball stands for Brand r . A ball is drawn at random. Suppose it was red (i.e., the consumer purchases brand r). The ball is replaced *and* d red balls are added to the urn. On the first purchase the probability of purchasing Brand r was $r/r + b$. Given the fact that Brand r has been purchased, the probability of purchasing Brand r on the next purchase has increased to $r + d/r + b + d$. Similarly, if Brand b were purchased at random on the first occasion, it would be replaced and c black balls would be added to the urn before the next draw. In this model the probability of the consumer purchasing a given brand on the next purchase is a function of the previous sequence of brand choices. Customer No. 2 is one example of the results of such a model.⁶

This definition of habitual behavior is an example of associative learning under conditions of reward such as those discussed by R. R. Bush and C. F. Mosteller in *Stochastic Models for Learning*.⁷

Ultimately, such a framework might prove useful as a basis for developing a theory of consumer purchasing behavior.

OPERATIONAL DEFINITIONS

Suppose the following is the chronological purchase record of a family consuming Brands X, Y, and Z for one year:

(XXX)	(YY)	(Z)	(X)	(Z)	(X)
1	2	3	4	5	6
(YYYYY)		(ZZ)	(X)	(Y)	
(7)		8	9	10	

The procedure for estimating the probabilities required to answer the first question is as follows:

1. Partition every family's purchase record into the runs⁸ of which it

⁶ For a discussion of the probability theory underlying models such as these see William Feller, *An Introduction to Probability Theory and Its Applications* (2d ed.; New York: John Wiley & Sons, 1957), 108-14.

⁷ (Chicago: John Wiley & Sons, Inc., 1955.) Those readers interested in the application of stochastic models to the behavioral sciences are advised to see T. W. Anderson, "Probability Models for Analyzing Time Changes in Attitudes," in Paul F. Lazarsfeld (ed.), *Mathematical Thinking in the Social Sciences* (2d ed.; Glencoe, Ill.: Free Press of Glencoe, Inc., 1955). See also a forthcoming book by James S. Coleman, of the Department of Social Relations at Johns Hopkins University.

⁸ A run is a set of consecutive purchases of the same brand.

is composed. In the above example there are ten runs, each in parentheses.

2. The first and last runs (runs 1 and 10 in our example) are omitted from the calculations because their exact length cannot be determined from the available data.

3. The remaining runs (2 through 9) are then classified by brand and by number of purchases to form a frequency distribution of the number of runs that are 1, 2, 3 . . . n purchases long for each brand.

4. Each of these distributions is then converted into a cumulative distribution of the number of runs that are more than n purchases long.

5. The estimated probability of staying with a given brand (P_s) on the $n + 1$ purchase, after having purchased the brand for n consecutive times is:

$$P_{s, n+1} = \frac{\text{Number of runs more than } n + 1 \text{ purchases long}}{\text{Number of runs more than } n \text{ purchases long}}$$

For example, suppose one wanted to calculate $P_{s,2}$ for a specific brand where 200 runs were longer than 1 purchase and 150 were longer than 2 purchases. This means that 50 of the runs were terminated after purchasing the brand once. Therefore, the probability of staying with the brand after making one purchase is $150/200$ or 0.75 .⁹

The procedures for answering the second question (i.e., given that a customer left a certain brand 1, 2, 3 . . . n purchases ago, what is the probability that he will return on the next purchase?) are similar to those for the first. The same family purchase record is used as an example. Suppose that the computation is done for Brand X. For the purpose of this computation the run length between runs of Brand X consists of all consecutive non-X purchases, not just consecutive purchases of the same brand. For example, the number of purchases made before returning to Brand X after leaving the first run is three. This "run" of non-X is actually composed of one run of Brand Y that is two purchases in length and one of Brand Z that is one purchase in length. Throughout this paper these runs are called "non-brand runs," in contrast to the term "brand runs" which is used to refer to the runs of a given brand. Using this new definition of a "non-X" run, the operational definition of the relationship is as follows:

1. Divide each family's purchase record into X and non-X (0) runs of which it is composed.

(XXX)	(000)	(X)	(0)	(X)
1	2	3	4	5
	(0000000)	(X)	(0)	
	6	7	8	

There are now 8 runs in our example.

⁹ It is shown later in this paper that this interpretation of probability may be fallacious.

2. The first and last runs, if they are non-X runs, are omitted from the calculations.
3. The remaining non-X runs are then classified by number of purchases to form a frequency distribution of the number of purchases 1, 2, 3 . . . n that it takes to return to Brand X after leaving it.
4. Each of these distributions is converted into a cumulative distribution of the number of runs more than n purchases long.
5. The probability of returning to Brand X (P_r) on the $n + 1$ purchase after having left the brand for n consecutive purchases is:

$$P_{r, n+1} = \frac{\text{Number of runs exactly } n \text{ purchases in length}}{\text{Number of runs more than } n - 1 \text{ purchases in length}}$$

For example, suppose one wanted to calculate $P_{r,3}$ for a given brand, knowing that there were 150 runs more than one purchase in length and fifty runs of two purchases in length. The probability of returning after two consecutive non-X purchases is $50/150$ or 0.33.

These procedures¹⁰ divide the family purchase patterns into runs of varying length for each brand.

When the data for all families purchasing a given brand are aggregated in a fashion aimed at answering the questions previously cited, the probability of repurchase does seem, at first, to be related to the number of preceding purchases (i.e., there appears to be habitual behavior). I shall give evidence that suggests that this relationship is spurious. At least for the majority of families purchasing a given brand one could just as well assume that their probability of purchasing the brand (as measured by the share of purchases they devoted to the brand during the period) remained constant.

ANALYSIS

Figure 1 presents the data for all coffee combined. At first the probability of purchasing a brand appears to increase as the number of previous consecutive purchases increases. Eventually, it seems to become constant. Similarly, the probability of returning to a brand after leaving it at first declines and then appears to become nearly constant. The increased variation in the probabilities as the number of consecutive purchases in a run increases is probably due primarily to the effect of sample size. For example, $P_{s,2}$ is based on a sample of 5,475 runs, while $P_{s,10}$ is based on a sample of only 65. But there is strong evidence that these relationships are largely spurious, as we shall see.

While there are variations in the rate at which the probabilities in-

¹⁰ The procedures outlined above weight families unequally and brand purchases equally. If a family switches brands frequently during the year it is represented by more runs than one which switches infrequently (assuming the total number of purchase decisions during the year for each of the two families is the same). Each brand choice regardless of the quantity involved is given the same weight in the computation.

crease or decrease for brand and non-brand runs respectively, the general shape of the relationships for individual brands appears to be about the same.

Table 2 presents the repeat and return purchase probabilities for each of five selected brands, while Table 3 presents the percentage of brand and non-brand runs greater than n purchases in length for each of the same five brands.

The data seems to suggest that the observed behavior with respect to patterns of brand choice is similar in appearance to what would be expected with associative learning under the conditions of reward.¹¹ The ob-

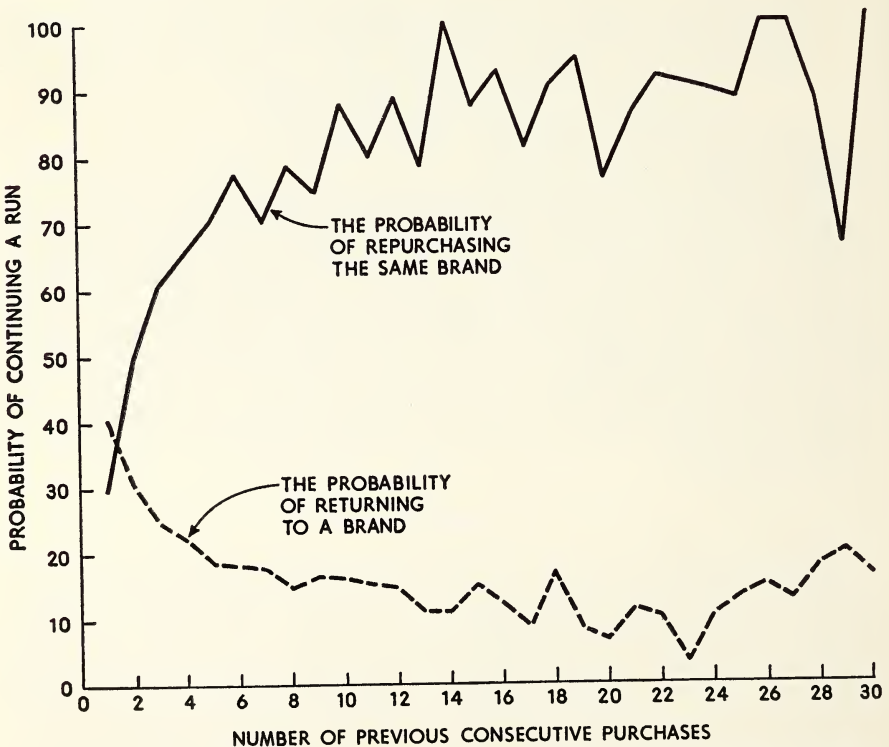


FIGURE 1. Probability of continuing a brand or nonbrand run for all coffee combined, after a run of n previous purchases.

served relationship between the probability of repurchasing a given brand and the number of previous consecutive purchases of the brand is based on an aggregation of the purchasing behavior of all of the families who bought the brand during the year under study. Such an aggregation can result in a spurious relationship (i.e., spurious contagion). It

¹¹ Though Kuehn (*op. cit.*) used a different method (he looked at all sequences of a given length regardless of their brand composition) his data also seems to suggest this interpretation.

TABLE 2
THE REPEAT AND RETURN PURCHASE PROBABILITIES FOR FIVE SELECTED BRANDS

Number of Previous Con- secutive Purchases	Hills Brothers		Manor House		Chase and Sanborn		Eight O'Clock (A & P)		Stewarts	
	Repeat	Return	Repeat	Return	Repeat	Return	Repeat	Return	Repeat	Return
1.....	32.4	35.3	29.1	36.2	30.8	35.0	39.3	53.2	32.9	52.6
2.....	49.4	36.3	47.6	35.0	45.6	24.8	51.5	33.6	54.7	29.6
3.....	62.6	24.4	47.4	16.8	54.2	26.7	68.7	29.1	65.5	47.4
4.....	67.5	20.4	59.5	22.7	59.0	21.7	67.4	28.6	68.4	35.0
5.....	76.9	15.0	81.8	19.6	87.0	20.8	64.5	17.5	92.3	23.1
6.....	75.0	16.8	66.7	23.0	80.0	10.0	65.0	18.2	58.3	50.0
7.....	76.7	21.3	66.7	17.5	68.8	18.1	76.9	7.4	71.4	40.0
8.....	87.0	18.9	100.0	8.5	81.8	11.9	80.0	16.0	80.0	0.0
9.....	70.0	13.3	100.0	14.0	66.7	21.2	75.0	19.1	100.0	33.3
10.....	78.6	23.1	100.0	5.4	100.0	14.6	50.0	17.7	100.0	0.0
11.....	90.9	20.0	87.5	11.4	83.3	22.9	66.7	21.4	75.0	0.0
12.....	100.0	18.8	85.7	6.5	100.0	18.5	50.0	9.1	100.0	0.0
13.....	70.0	3.9	66.7	10.3	100.0	22.7	100.0	0.0	66.7	0.0
14.....	100.0	20.0	100.0	7.7	100.0	11.8	100.0	10.0	100.0	50.0
15.....	100.0	5.0	100.0	8.3	80.0	26.7	100.0	11.1	50.0	0.0
16.....	100.0	15.8	100.0	22.7	100.0	18.2	100.0	12.5	100.0	0.0
Number of runs.....	(793)	(583)	(513)	(345)	(564)	(372)	(331)	(254)	(161)	(114)

TABLE 3
PERCENTAGE OF BRAND AND NONBRAND RUNS GREATER THAN
 n PURCHASES IN LENGTH FOR FIVE SELECTED BRANDS

Number of Previous Con- secutive Purchases	Hills Brothers		Manor House		Chase and Sanborn		Eight O'Clock		Stewarts	
	Non- Brand	non- brand	Non- Brand	non- brand	Non- Brand	non- brand	Non- Brand	non- brand	Non- Brand	non- brand
1.....	32.4	59.5	30.8	63.8	29.1	65.1	39.3	46.9	32.9	47.4
2.....	16.0	37.9	14.0	41.5	13.8	47.3	20.2	31.1	18.0	33.3
3.....	10.0	28.6	7.6	34.5	6.6	34.7	13.9	22.1	11.8	17.5
4.....	6.8	22.8	4.5	26.7	3.9	27.2	9.4	15.8	8.1	11.4
5.....	5.2	19.4	3.9	21.5	3.2	23.4	6.0	13.0	7.5	8.8
6.....	3.9	16.1	3.1	16.5	2.1	19.4	3.9	10.6	4.4	4.4
7.....	3.0	12.7	2.1	13.6	1.8	15.9	3.0	9.8	3.1	2.6
8.....	2.6	10.3	1.8	12.5	1.8	14.0	2.4	8.3	2.5	2.6
9.....	1.8	8.9	1.2	10.7	1.8	11.0	1.8	6.7	2.5	1.8
10.....	1.4	6.8	1.2	10.1	1.8	9.4	0.9	5.5	2.5	1.8
11.....	1.3	5.5	1.0	9.0	1.2	7.3	0.6	4.3	1.9	1.8
12.....	1.3	4.5	1.0	8.4	1.1	5.9	0.3	3.9	1.9	1.8
13.....	0.9	4.3	1.0	7.5	0.7	4.6	0.3	3.9	1.2	1.8
14.....	0.9	3.4	1.0	7.0	0.7	4.0	0.3	3.5	1.2	0.9
15.....	0.9	3.3	0.8	6.4	0.7	3.0	0.3	3.2	0.6	0.9
16.....	0.9	2.7	0.8	4.9	0.7	2.4	0.3	2.8	0.6	0.9

may be that for individual families the probability of purchasing a certain brand remains constant (no habitual behavior) but that aggregating families with different, *though constant*, purchase probabilities creates what appears to be a dependence.

SPURIOUS CONTAGION

The data to be presented next strongly suggest that the relationship between $P_{s,n+1}$ and n and $P_{r,n+1}$ and n are spurious. The discussion is divided into two parts: first, a slight digression aimed at clearly stating what I mean by "spurious contagion." Second, presentation of the analysis.

*An illustration of spurious contagion.*¹² The conceptual framework previously presented looks upon the purchase of a brand as a game of chance. Now the chance of purchasing a given brand may vary from person to person. Suppose there are just two types of people and that their numbers in the total population stand in the ratio 1:5. Consider that Urn I contains r_1 red and b_1 black balls and Urn II contains r_2 red and b_2 black balls.

Suppose a red ball signifies the purchase of Brand r and a black ball the purchase of Brand b . Each urn represents a particular group of customers with a constant probability of purchase for each brand. The choice of brand for a given group is determined by random drawing with replacement.

Suppose the two groups of people are combined in the analysis. The probability of a red ball being drawn on the first occasion is equal to the weighted average of the probabilities for the two groups of people (urns). The weights correspond to the relative importance of the two groups in the population (i.e., $1/6$ for Urn I and $5/6$ for Urn II). The probability of drawing a red ball $P(R)$ from a population consisting of both of the urns combined is:

$$P(R) = \frac{1}{6} \cdot \frac{r_1}{b_1 + r_1} + \frac{5}{6} \cdot \frac{r_2}{b_2 + r_2}$$

and the probability of a sequence red, red

$$P(RR) = \frac{1}{6} \cdot \left(\frac{r_1}{b_1 + r_1} \right)^2 + \frac{5}{6} \cdot \left(\frac{r_2}{b_2 + r_2} \right)^2$$

Suppose a customer purchases Brand r . What is the probability of purchasing Brand r on the next purchase? In other words, given that the first drawing resulted in red, we ask for the probability of a sequence red, red. This is clearly the ratio $P(RR)/P(R)$ and is different from $P(R)$. For the sake of illustration suppose that $r_1/(b_1 + r_1) = 0.6$ and $r_2/(b_2 + r_2) = 0.06$. The probability of red at any drawing is 0.15, but if the first drawing resulted in red, the chance of red on the next drawing is 0.42. Note that our model involves no aftereffect [i.e., that the probability of purchasing Brand r for a given group of people (or customer) remains constant despite their past experience in the total population]

¹² This illustration is adopted from one on accident proneness presented by Feller, *op. cit.*, p. 109 and pp. 111-12.

and yet the purchase of Brand r for a person chosen at random increases the odds that *this same person* will purchase Brand r on the next occasion. We have here an effect of sampling; the purchase of a brand does not have a real effect, but it is an indication that the person chosen at random has a high proneness to the brand.

The following analysis is concerned with the question: Is what appears to be a dependence for all families combined, actually the result of combining families with different, *though constant*, probabilities of purchasing Brand A? In other words, can the observed aggregate relationship be taken as evidence of some process similar to associative learning under the conditions of reward. Or, alternatively, is it simply an example of spurious contagion resulting from the aggregation of many customers with different and constant probabilities of purchasing Brand A.

Evidence of Spurious Contagion. Two types of evidence are presented that strongly suggest that a large part of the observed "learning" effect is actually due to differences in the purchase probabilities among the families. The first kind of evidence uses a Monte Carlo technique to replicate the observed purchasing behavior by family for a given brand assuming that a given family's probability of purchasing the brand is constant. The second examines runs by families to see what proportion of the families purchasing Brand A exhibit a constant probability of purchase for the brand during the period.

The Monte Carlo Approach. The empirical relationship between $P_{s,n+1}$ and n reflects the effects of both differences among families and dependency,¹³ between the consecutive purchases of individual families. Suppose we assumed independence by families, use the observed family purchase probabilities by brands, and generate a runs distribution. Comparison of this generated distribution with the actual distribution shows

¹³ The concept of dependence is not (as it is used here) necessarily synonymous with that of habit (that is, the effect of purchases made in the immediate past on the probability distribution with respect to the brand chosen on the next purchase). For example, suppose a family has a brand share for Brand A of 0.30 for the year. Actually the probability of buying A might have been 0.90 for the first few months and 0.10 for the remainder of the time. This could produce fewer runs than might be expected on the basis of a 0.30 probability (that is, a slower rate of brand-switching). This clustering, however, would not be produced by a dependency between successive purchases (that is, the effect of habit) but by an abrupt shift in the probability (which otherwise remained constant) during the year. Alternatively, suppose two members of the same family, with equal rates of consumption, purchased different brands. Their purchase pattern might look like this: ABABAB, with the husband consuming Brand A and the wife Brand B. The probability of purchase for a given member of the family is one. In a case such as this, the number of observed runs is greater than expected by chance (that is, a faster rate of brand-switching than that based on chance expectation). In other words, the observed pattern for the family suggests some degree of association between the consecutive purchases. Actually, there is no association, just a variation in the brand consumed by different members of the family. In either event, it is clear that dependence as the term is used here can be the result of factors other than the association between consecutive purchases, such as a shift in the customer's probability distribution of brand choice or different probabilities of buying the brands by members of the family.

how much of the dependence can be explained by heterogeneity among families.

To accomplish this, an attempt was made to estimate what the relationship between $P_{s,n+1}$ and n would have been if the only factor accounting for it is differences between families with respect to their probability of purchasing the brand. By way of describing the procedures used to estimate this relationship, a detailed example of the steps involved is presented for Hills Brothers coffee.

In making the estimate two pieces of information are used for each family purchasing Hills Brothers coffee: the number of purchase decisions the family made during the period and the share of these purchase decisions that went to Hills. In effect, this procedure attempts to answer the following question: Conditional upon the purchase decision share and total purchases of a given family, what would the family's purchase record look like if the successive brand choices were independent of each other?

Using Rand's table of *A Million Random Digits with 100,000 Normal Deviates* a purchase pattern was generated for each family in the following fashion: Suppose a given family bought coffee ten times during the year and Hills four times. The occurrence of digits one to four in the table were scored as a Hills Brothers purchase, while six to zero were counted as non-Hills purchases.

The procedures for estimating the relationship between $P_{s,n+1}$ and n from these "generated" chronological purchase records were exactly the same as that for the actual data.

Relationships such as these were generated for four of the nineteen brands included in the original tabulation. These were A & P, Eight O'Clock, Hills Brothers, and Stewarts.

Figure 2 shows the actual and the generated curves for Hills Brothers, and Table 4 presents the actual and generated probabilities for the other three brands. The similarity between the actual and generated curves for each of the four cases is striking. The two brands for which the difference between the actual and generated curves appears to be the greatest, A & P and Stewarts, are based upon the smallest sample sizes. This suggests that the principal factor generating the observed relationships is the difference among families with respect to their brand purchase probabilities and not dependence between the consecutive purchases of individual families. Or, to put it another way, as the number of previous consecutive purchases goes from 1 to n , a higher proportion of the runs are accounted for by families with high purchase probabilities (i.e., greater than 0.75). The data in Table 5 for Hills Brothers coffee indicate that a positive relation does exist. Of the 524 runs one purchase in length, 1.7 per cent came from families who bought Hills more than 75 per cent of the time while 36.5 per cent of the runs longer than four purchases in length came from the same group.

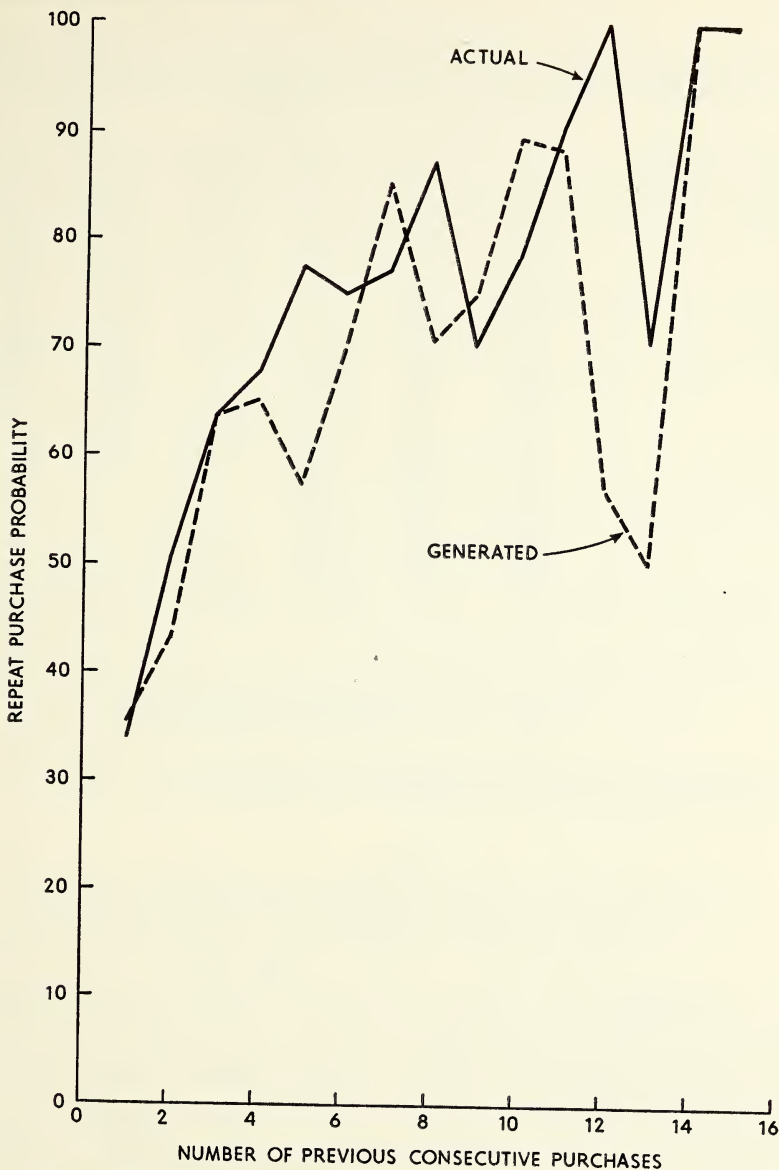


FIGURE 2. Actual and generated relationship between probability of repurchasing a brand on next purchase ($n+1$) and number of previous consecutive purchases (n) for Hills Brothers coffee.

These results suggest that a large part of the dependent relationship between $P_{r,n+1}$ and $P_{s,n+1}$ and n is spurious. The model is that used to replicate the brand purchasing behavior for a given brand using the random numbers table is based on the assumption that the purchase probability for a given family remained constant (i.e., that there is no habitual

TABLE 4

ACTUAL AND GENERATED REPEAT PURCHASE PROBABILITIES FOR THREE SELECTED BRANDS

Number of Previous Purchases	Eight O'Clock		Stewarts		A & P	
	Actual	Generated	Actual	Generated	Actual	Generated
1.....	39.3	37.2	32.9	35.1	26.4	21.1
2.....	51.5	50.0	54.7	55.9	48.5	40.6
3.....	68.7	67.1	65.5	69.7	56.3	61.5
4.....	67.4	72.3	68.4	73.9	66.7	87.5
5.....	64.5	70.6	92.3	82.4	83.3	57.1
6.....	65.0	83.3	58.3	78.6	60.0	75.0
7.....	76.9	60.0	71.4	72.7	66.7	100.0
8.....	80.0	58.3	80.0	75.0	100.0	33.3
9.....	75.0	85.7	100.0	66.7	100.0	0.0*
10.....	50.0	83.3	100.0	50.0	50.0	
11.....	66.7	100.0	75.0	60.0	100.0	
12.....	50.0	100.0	100.0	0.0*	0.0*	
13.....	100.0	100.0	66.7			
14.....	100.0	40.0	100.0			
15.....	100.0	50.0	50.0			
16.....	100.0	100.0	100.0			
Total number of runs..	(331)	(376)	(161)	(168)	(125)	(152)

* The probability of staying with a brand was defined as:

$$P_{s, n+1} = \frac{\text{Number of runs more than } n + 1 \text{ purchases long}}{\text{Number of runs more than } n \text{ purchases long}}$$

A zero probability arises when all runs that are more than n purchases in length terminate on the $n + 1$ purchase. This in turn means that there were no runs longer than $N + 1$ purchases observed for that brand, and therefore there can be no repeat purchase probability estimates for the $n + 2$ purchase.

TABLE 5

PERCENTAGE OF RUNS FOR HILLS BROTHERS COFFEE OF LENGTH n PRODUCED BY FAMILIES WITH SELECTED PURCHASE DECISION SHARES

Purchase Decision Share	Run Length				
	1	2	3	4	5 and Over
Greater than 50 percent but equal to or less than 100 percent....	13.93	35.86	48.94	60.00	84.61
Greater than 75 percent but equal to or less than 100 percent....	1.72	4.14	8.51	12.00	36.54
Total number of runs.....	(524)	(145)	(17)	(25)	(52)

behavior). This assumption seems adequate to generate a relationship that looks like the one based on the actual data. These results cast suspicion on the use of a "learning" model to describe the observations.

Aggregate models of this nature may, however, be useful in other contexts. For example, it may be that this aggregate model will lead to useful short-run predictions of market share even though it provides little insight into loyalty-prone behavior.

The fact that the assumption of a constant purchase probability for a

given family for a given brand does appear to account for the observed patterns of brand choice for the family raises still another question. Can a simple probability model be found that accounts for the level of the purchase probability for each family?

The model based on constant probabilities for a given family and varying probabilities from family to family provides no mechanism for explaining why Brand X has a 20 per cent share of one family's purchases and 100 per cent share of another's.

The Distribution of Runs Approach. The model that is presented provides an answer to the following question: To what extent does the observed pattern of brand choices over time for a given family support the hypothesis that the probability of purchasing a certain brand is constant during the period under study?

Data for Hills Brothers and Chase and Sanborn coffee are presented. The model compares the actual distribution of runs with the distribution of runs that is expected by chance conditional upon a certain probability of purchasing the brand (operationally defined as a family's purchase decision share) and a certain number of purchase decisions during the year. For example, suppose a family has a probability of .75 of purchasing Brand A during a given year. If the family made 100 purchases during that year, the expected number of runs, assuming all non-Brand A runs are grouped as one brand, would be:¹⁴

$$M = \frac{2n_1n_2}{n} + 1 = \frac{2(75)(25)}{100} + 1 = 38.5$$

where

n_1 is the number of Brand A purchases

n_2 is the number of non-Brand A purchases

n equals n_1 plus n_2

One can compare the expected number of runs, 38.5, to the actual number. The greater the clustering of the runs of Brand A (i.e., the fewer the runs) conditional upon a given probability of purchasing Brand A, the stronger the evidence that some change in probability occurred during the period. The standard normal distribution can be used as a test of the extent to which the actual clustering differs from what one expects by chance. If, in our example, the actual number of runs is 30, then the normal deviate (K) is computed in the following fashion, where:

$$K = \frac{r + \frac{1}{2} - M}{\sigma_r}$$

$$\sigma_r = \sqrt{\frac{2n_1n_2(2n_1n_2 - n)}{n^2(n - 1)}}$$

¹⁴ W. Allen Wallis and Harry V. Roberts, *Statistics: A New Approach* (Glencoe, Ill.: Free Press of Glencoe, Inc., 1956), pp. 569-70.

$$K = \frac{30 + \frac{1}{2} - 38.5}{\sqrt{\frac{2(75)(25)[2(75)(25) - 100]}{(100)^2(100 - 1)}}} = -2.16$$

These measures cannot be interpreted as a reflection of the effect of any specific variable on brand choice. For example, clustering over and above what is expected on the basis of chance may be due to habitual purchasing. Or, there might have been a change in the availability of one of the brands the family tends to purchase.

At the other extreme, more runs than are expected by chance might suggest that the family alternates between a couple of favorites or that different members of the family prefer different brands.

A second limitation of this procedure is that the data for a majority of the families could not be used in the process of estimation. The statistical model cannot detect association between consecutive purchases for very small sample sizes and for situations in which the probability of buying a single brand is quite high. For the purpose of this initial exploration, families with fewer than twenty purchase decisions during the year are excluded. In addition, those families that have over twenty purchases but made fewer than four purchases either for the brand being tested or for all other brands purchased are also excluded (115 families for Hills and 78 for Chase and Sanborn). Of the original 362 families that purchased Hills and the 219 families that purchased Chase and Sanborn, only 71 and 53, respectively, remained after applying the above criteria.

For each brand the normal deviate for each family is arrayed and converted into a percentage distribution, which consists of the percentage of families with a normal deviate of less than -5 , greater than or equal to -5 but less than -4 , . . . , greater than zero but less than or equal to 1 , . . . , and greater than 5 . There are no zero values among the actual deviates.

Figure 3 presents both the actual percentage distribution of normal deviates for Hills Brothers and Chase and Sanborn and the standard normal distribution. The standard normal distribution presents the expected result if the probability for each family remained constant during the period. Both Hills Brothers and Chase and Sanborn have a considerably higher percentage of negative normal deviates in the tail than is expected on the basis of chance assuming a constant probability of purchase for individual families. In other words, there are a number of runs that are longer than expected by chance. Also, however, a large proportion of families purchasing both brands behave as if they had stable purchase probabilities. If each family showed independence and a constant probability of purchase, we would expect 68 per cent of the families to have deviates between -1 and 1 , whereas 51 per cent of the Hills Brothers families and 58 per cent of the Chase and Sanborn families fall in the

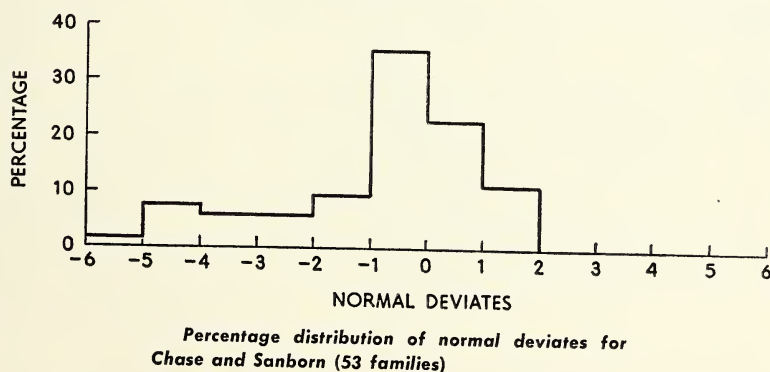
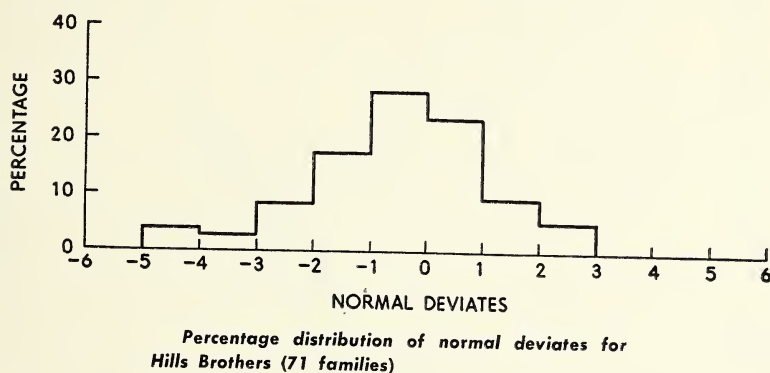
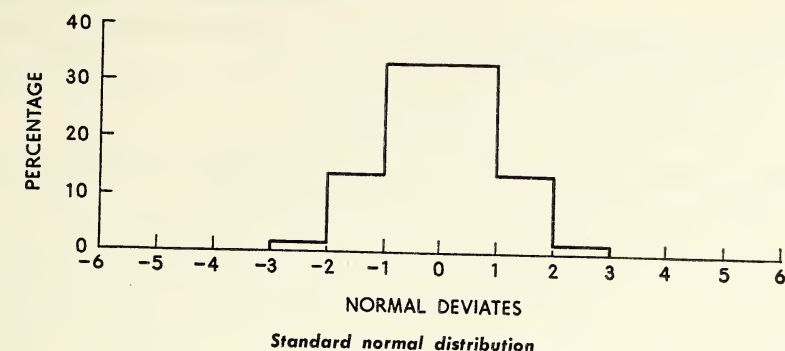


FIGURE 3. Actual percentage distribution of normal deviates for selected Hills Brothers and Chase and Sanborn families and standard normal distribution.

interval.

An alternative way of looking at this behavior is to look at less than cumulative distributions of the normal deviates by brand on arithmetic probability paper (Fig. 4). The broken lines are the actual cumulative distributions. Dots are used as symbols for Hills Brothers and x's for Chase and Sanborn. The solid line represents the expected cumulative

distribution assuming that the probability of purchase for each family remains constant during the period.

In both cases, the mean of the actual distributions is negative, which is

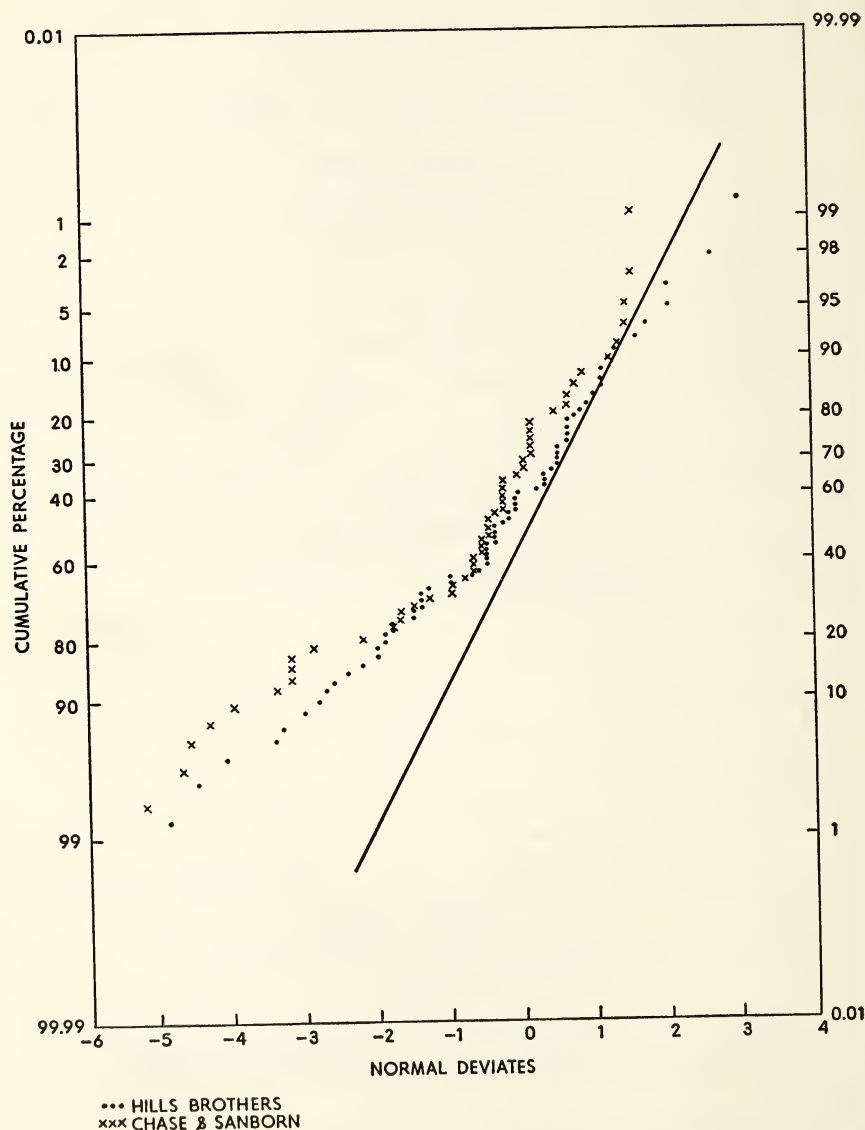


FIGURE 4. Less-than-cumulative distribution of normal deviates for selected Hills Brothers and Chase and Sanborn families.

what one would expect if there are non-constant probabilities of purchase for individual families. The variance of the actual distribution is greater than one. This is implied by the fact that the slope of the actual cumula-

tive is less than the slope of the standard normal distribution, which has a variance of one.

CONCLUSIONS

The analysis of repeat purchase probabilities by brand seems to suggest that: (1) The model of constant probability of purchase fails in the tails of the distribution. There are too many families with an excess of long runs. On the other hand, there are many families whose behavior is consistent with the hypothesis of a constant probability of purchase during the period. (2) A minority of the families appear to have a non-constant probability of purchasing the brand. Of course, the data say nothing about what factors caused the probabilities to shift.

For the purpose of separating out these effects further research is needed in at least two directions: (1) Toward the construction of considerably more complex models that take into account the effects of a brand's relative position with respect to such variables as price, promotion, and distribution as well as sequences of past purchases; (2) Toward the study of market situations where abrupt changes, such as in the introduction of a new brand, have occurred. Without some abrupt change it may be that many families have moved to equilibrium position with respect to the distribution of their choices among brands. A learning model may be relevant for explaining how they arrived at equilibrium. But, once there, one of the sequences of choices for individual families might appear to be consistent with a simple model that assumes no learning.

While these findings shed light on one point, they raise questions with respect to another which merits further investigation: Can a model be found that will account for the variation in the probability of purchasing a given brand from family to family?

It is hoped that this analysis, in addition to providing some information about the processes underlying the observed behavior, will lead to a more critical evaluation of the use of probability models in developing a theory of brand choice.

Consumer Brand Choice—A Learning Process?*

ALFRED A. KUEHN†

THE PHENOMENON OF CONSUMER BRAND SHIFTING IS A CENTRAL ELEMENT underlying the dynamics of the marketplace. To understand and describe market trends adequately, we must first establish the nature of the influences on consumer choice with respect to products and brands. Research directed at establishing the conditions under which consumers will shift from one brand to another offers hope of providing a framework within which to evaluate the influence of price, advertising, distribution and shelf space, and various types of sales promotion.

What do we know about brand choice? What behavioral mechanisms appear to underlie this phenomenon? Is such behavior habitual? Is learning involved? Does repeated purchasing of a brand reinforce the brand choice response? What is the relationship between consumer purchase frequencies and brand shifting behavior? These questions will be discussed in the light of available empirical data and a model which appears to describe them.

A MODEL OF CONSUMER BRAND SHIFTING

A model equivalent to a generalized form of the Estes¹ and Bush-Mosteller² stochastic (probabilistic) learning models appears to describe

* The research underlying this paper has been supported through grants from the Graduate School of Industrial Administration and the Market Research Corporation of America. The paper is based in part upon a series of lectures presented at the Ford Foundation Faculty Seminar in Marketing conducted by the University of Chicago at Williamstown, Massachusetts, August, 1961. It is tentatively scheduled for publication in the December, 1962, issue of the *Journal of Advertising Research*.

† Carnegie Institute of Technology.

¹ William K. Estes, "Individual Behavior in Uncertain Situations: An Interpretation in Terms of Statistical Association Theory," in Thrall, R. M., C. H. Coombs, and R. L. Davis (eds.), *Decision Processes* (New York: John Wiley & Sons, Inc., 1954).

² Robert R. Bush and Frederick Mosteller, *Stochastic Models for Learning* (New York: John Wiley & Sons, Inc., 1955).

consumer brand shifting quite well. To illustrate how this brand shifting model describes changes in the consumer's probability of purchasing any given brand as a result of his purchases of that brand (for example, Brand A) and competing brands (for example, Brand X), let us examine the effect of the four-purchase sequence XAAX upon a consumer with initial probability $P_{A,1}$ by referring to Figure 1.

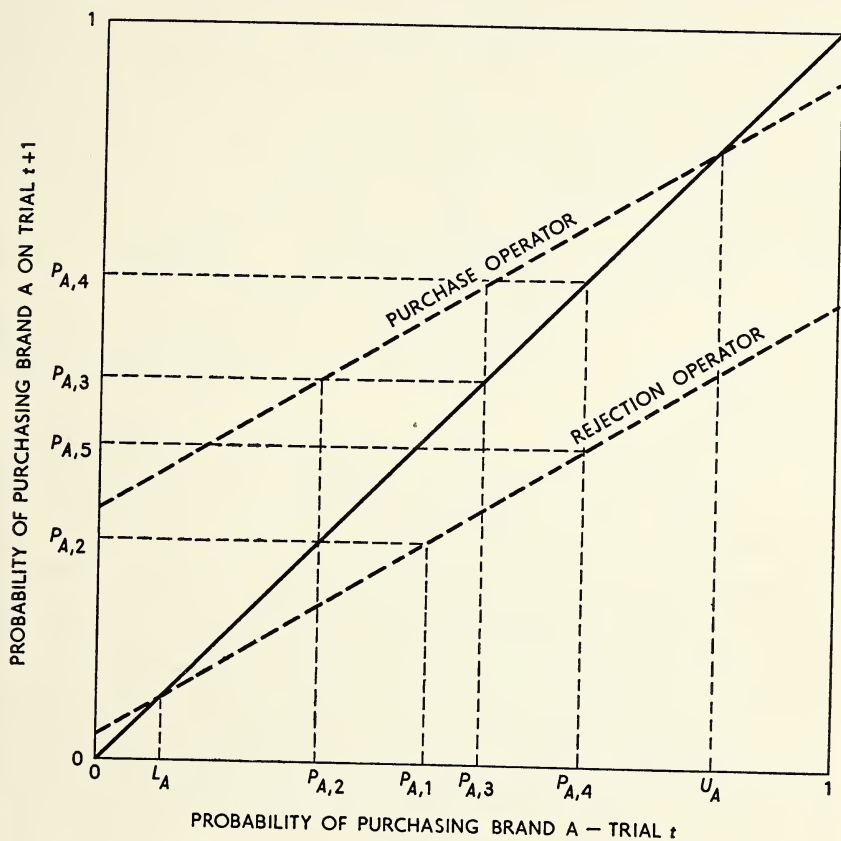


FIGURE 1. Stochastic (probabilistic) brand shifting model.

The model is described or defined in terms of four parameters, namely, the intercepts and slopes of the two lines referred to in Figure 1 as the Purchase Operator and the Rejection Operator. If the brand in question is purchased by the consumer on a given buying occasion, the consumer's probability of again buying the same brand the next time that type product is purchased is read from the Purchase Operator. If the brand is rejected by the consumer on a given buying occasion, the consumer's probability of buying that brand when he next buys that type product is read from the Rejection Operator. Thus, note in Figure 1 that our hypothetical consumer begins on trial 1 with the probability $P_{A,1}$ of

buying Brand A. The consumer chooses some other brand (X) on trial 1, however, and thus his probability of buying Brand A on trial 2 ($P_{A,2}$) is obtained from the Rejection Operator, resulting in a slight reduction in the probability of purchasing A on the next trial. On trial 2, however, the consumer does purchase Brand A and thus increases the likelihood of his again buying the brand on the next occasion (trial 3) to $P_{A,3}$. Continuing in this fashion, the consumer again buys A on trial 3, thereby increasing his probability of purchasing Brand A on trial 4 to $P_{A,4}$. He again rejects A on trial 4, however, decreasing his probability of buying A on trial 5 to $P_{A,5}$.

Two characteristics of the model should be noted: (1) The probability $P_{A,t}$ approaches but never exceeds the upper limit U_A with repeated purchasing of the brand, and (2) the probability $P_{A,t}$ approaches but never drops below the lower limit L_A with continued rejection of the brand. Using Bush and Mosteller's terminology, this would be referred to as an incomplete learning, incomplete extinction model insofar as U_A is less than 1 and L_A is greater than 0. This is equivalent to saying that consumers will generally not develop such strong brand loyalties (or buying habits) as to insure either the rejection or purchase of a given brand.

It should also be pointed out that the Purchase and Rejection Operators are functions of the time elapsed between the consumer's t^{th} and $t + 1^{\text{st}}$ purchases and the merchandising activities of competitors. The time effect can be illustrated by the three sets of operators shown for high, medium, and very low frequency purchasers of a rapidly consumed, nondurable consumer product (see Figure 2). Note that the slopes of the Purchase and Rejection Operators decrease and that the upper and lower limits approach each other as the time between purchases increases.

At the one limit (Δ time between purchases approaching 0) the Purchase and Rejection Operators approach the diagonal, L approaches 0, and U approaches 1. At the other limit (Δ time between purchases approaching ∞), L and U approach each other and the Purchase and Rejection Operators approach a slope equal to 0.

The main problem that remains in making use of the model is then the estimation of the four parameters defining the Purchase and Rejection Operators as a function of the time between purchases. If this could be done a priori, the model might be of value to marketing management for use in forecasting. At present, however, the model's primary use is in evaluating the effects of past and current competitive marketing activity. Thus, the parameters of the model are estimated for short time periods and related to the actions of all competitors in the market. Since the path of aggregate consumer purchasing behavior could be established for any given set of parameter values, it follows that the parameter estimates obtained from fitting the model can provide a means for evaluating the

influence of the market conditions prevailing during the period in which the sequential purchase data are collected.

An efficient method has been developed to estimate these brand shifting parameters (maximum likelihood estimates) on the basis of sequences of two to four purchases. This makes it feasible to relate this model to consumer purchasing behavior observed during relatively short periods of time. This is a must if the technique is to be useful, since merchandising conditions do not remain constant for long periods of time—products are modified, advertising themes and budgets are altered, special promotions are generally temporary in nature, and price levels may change from time to time. The technique used to estimate the brand shifting parameters will be outlined in the near future as a working paper in the Carnegie Tech (GSIA) Research in Marketing Project series.

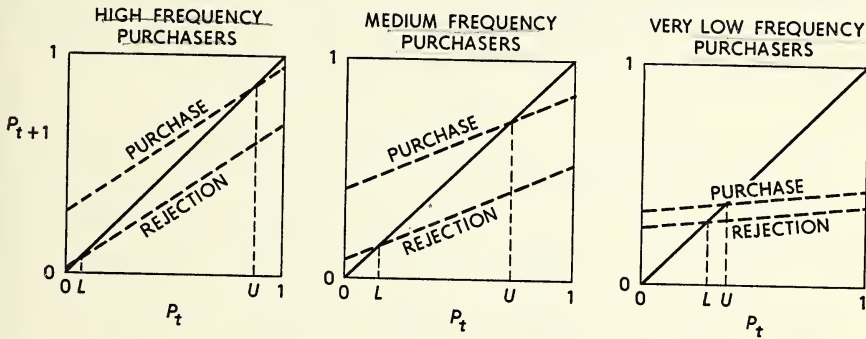


FIGURE 2. Effect of time between purchases upon purchase and rejection operators.

The Bush-Mosteller approach to estimating the parameters of their stochastic learning model cannot, in its current state of development, be applied to the brand shifting model since (1) techniques have not been developed to estimate simultaneously the four basic parameters of the model, and (2) the methods outlined require a long history or record of trials (and, therefore, data collected over a long period of time during which there is stability in merchandising activity) from which to develop parameter estimates.

EMPIRICAL BRAND SHIFTING RESEARCH

What evidence is there in support of the model? Three types of empirical studies have led to the formulation and continued development of the above model:³

1. Analysis of 3, 4, 5, and 6 purchase sequences of consumer brand purchases.

³ The first two of the following three studies are reported in detail in Alfred A. Kuehn, "An Analysis of the Dynamics of Consumer Behavior and Its Implications for Marketing Management" (unpublished Ph.D. dissertation, Graduate School of Industrial Administration, Carnegie Institute of Technology, 1958).

2. Analysis of effects of time between consumer purchases upon a consumer's probability of purchasing individual brands of product.
3. Simulation of consumer brand choice behavior.

Each of these three studies is discussed briefly below.

Analysis of Brand Purchase Sequences

Sequential purchase data can provide some insight into consumer brand switching. The data analyzed below represent the frozen orange juice purchases of approximately 600 Chicago families in the three years 1950 to 1952. More than 15,000 individual purchases of frozen orange juice were collected in monthly diaries by the *Chicago Tribune* Consumer Panel during this period. Data were analyzed as sequences of five purchases by means of a factorial analysis to determine the influence of the consumer's first four brand choices within each sequence upon his choice of a brand on the next (fifth in the sequence) buying occasion. The data and analysis prepared for the Snow Crop brand are summarized in Table 1.

TABLE 1
COMPARISON OF OBSERVED AND PREDICTED PROBABILITY OF PURCHASING SNOW
CROP GIVEN THE FOUR PREVIOUS BRAND PURCHASES

<i>Previous Purchase Pattern (1)</i>	<i>Sample Size (2)</i>	<i>Observed Probability of Purchase (3)</i>	<i>Predicted Probability of Purchase (4)</i>	<i>Deviation of Predictions (5)</i>
SSSS.....	1,047	0.806	0.832	+0.026
OSSS.....	277	0.690	0.691	+0.001
SOSS.....	206	0.665	0.705	+0.040
SSOS.....	222	0.595	0.634	+0.039
SSSO.....	296	0.486	0.511	+0.025
OOSS.....	248	0.552	0.564	+0.012
SOOS.....	138	0.565	0.507	-0.058
OSOS.....	149	0.497	0.493	-0.004
SOSO.....	163	0.405	0.384	-0.021
OSSO.....	181	0.414	0.370	-0.044
SSOO.....	256	0.305	0.313	+0.008
OOSOS.....	500	0.330	0.366	+0.033
OOSO.....	404	0.191	0.243	+0.052
OSOO.....	433	0.129	0.172	+0.043
SOOO.....	557	0.154	0.186	+0.032
OOOO.....	8,442	0.048	0.045	-0.003

In column 1, the letter "S" is used to represent a purchase of the Snow Crop brand, the letter "O" to represent the purchase of any brand of frozen orange juice *other* than Snow Crop. Thus the sequence SSSS indicates a sequence of four purchases of Snow Crop. The sequences OSSS represents one purchase of some brand other than Snow Crop followed by three purchases of Snow Crop.

Column 2 tabulates the sample sizes from which the observed and predicted probabilities of purchasing Snow Crop on the subsequent buying occasion (fifth purchase in the sequence) were calculated.

Column 3 is computed on the basis of the observed frequencies of the five-purchase sequences. Thus, there were 296 sequences exhibiting the pattern SSSO in the first four positions of the sequence. Snow Crop was purchased on the fifth buying occasion in 144 of these sequences. The best estimate of the observed probability of buying Snow Crop given the past purchase record of SSSO is therefore $144/296 = 0.486$.

The predicted column is based upon the results of the previously referred to factorial analysis of past purchase effects. Each of the four past brand purchases were examined with respect to their individual (primary) effects and the effects of their interactions with each other. The individual effects of the past four purchase positions were highly significant but the interaction effects were not significantly different from 0 at the 5 percent level of significance (that is, there was greater than 5 percent probability of results as extreme as those observed arising by chance if there were in fact no interaction effects).

There is close agreement between the observed and predicted probabilities in view of the limited sample size. There appear, however, to be systematic deviations on the high side when Snow Crop is purchased either one or three times (also predictions are generally low given two purchases) during the last four buying occasions. Subsequent analysis indicated that these systematic deviations were reduced or eliminated when a record of the fifth past brand purchase was included in the analysis.

Casual inspection of Table 1 suggests that the most recent purchase of the consumer is not the only one influencing his brand choice. This finding raises some question about the uses currently being made of purchase-to-purchase Markov Chain Analyses which assume that only the most recent purchase of the consumer is influential. The analysis of "primary" effects referred to above showed that the purchase of Snow Crop on the most recent buying occasion added 0.321 to the probability of the consumer buying Snow Crop on his next purchase. Similarly, the second most recent purchase added 0.198, the third 0.127 and the fourth 0.141.⁴

Note that the first three purchase effects decline roughly exponentially. That is, the ratio of the importance of the first purchase to that of the second is approximately equal to the ratio of the second to the third. The fourth, however, increases rather than decreases! This reversal has been traced to the fact that past purchases beyond the fourth most recent purchase were excluded from the analysis. The increased importance attached to the fourth most recent purchase for purposes of predic-

⁴ To illustrate the computation of the predicted probabilities in column 4, Table 1, the probability of a Snow Crop purchase given the history SOOO is 0.045 (the probability of purchase given OOOO) plus 0.141 or 0.186, the probability given SOOS is $0.045 + 0.141 + 0.321 = 0.507$ and the predicted probability given OSSS is $0.045 + 0.127 + 0.198 + 0.321 = 0.691$.

tion reflects its high correlation with the fifth and earlier past purchases not incorporated in the study. When these same data were re-analyzed using six-purchase sequences, the exponential relationship of declining primary purchase effects fit the first through fourth past purchases. As would be expected, however, the fifth past purchase effect was larger than the fourth because of its higher correlation with the consumer's sixth and even earlier past purchases.

Observation of the exponentially declining effects of past purchases led to the testing of the brand shifting model outlined in Figure 1 since that model has the characteristic of weighting the influence of past brand choices exponentially when the slopes of the Purchase and Rejection operators are identical. Subsequent research with products other than frozen orange juice has tended to confirm the exponential weighting of past brand purchases by consumers for predictive purposes. The exponential weights vary substantially, however, among product classes. Products such as toilet soap, cereals, and toothpaste were found to have substantially lower rates of decline in weights as one goes back into the purchase history as a result of the tendency of purchasing families to use some mix of brands on a routine basis to satisfy different uses, desires for variety, and differences in preference of individual family members. To be sure, this brand-mix effect is operative even in the case of frozen orange juice but for quite a different reason. Many families use a mix of brands of frozen orange juice because of the lack of availability of individual brands of product in all of the stores among which the consumer shifts in the course of his week-to-week shopping trips.

Effect of Consumer Purchase Frequencies

Let us consider the effect of time between purchases upon the consumer's probability of repurchasing the same brand. In Figure 3 we observe the probability of a consumer's buying the same brand on two consecutive purchases of the product decreasing to the share of market of the brand as time between purchases increases. Whenever a great amount of time has elapsed since the consumer's last purchase of the product, the brand he last bought has little influence upon his choice of a brand—the probability of his buying any given brand in this case is approximately equal to the share of market of that brand. It should be noted that the probability of repurchase decreases at a constant rate with the passing of time; this characteristic, which we shall refer to as the “time rate of decay of purchase probability,” is significant since it provides a simple framework within which to incorporate the effects of time into a procedure for forecasting consumer purchase probabilities.

Let us now expand our view of the effects of time upon repurchase probability in terms of the time period required for the consumer to make N individual purchases of frozen orange juice concentrate. Note that the curve in Figure 4 labeled $N = 1$ is the same curve as in Fig-

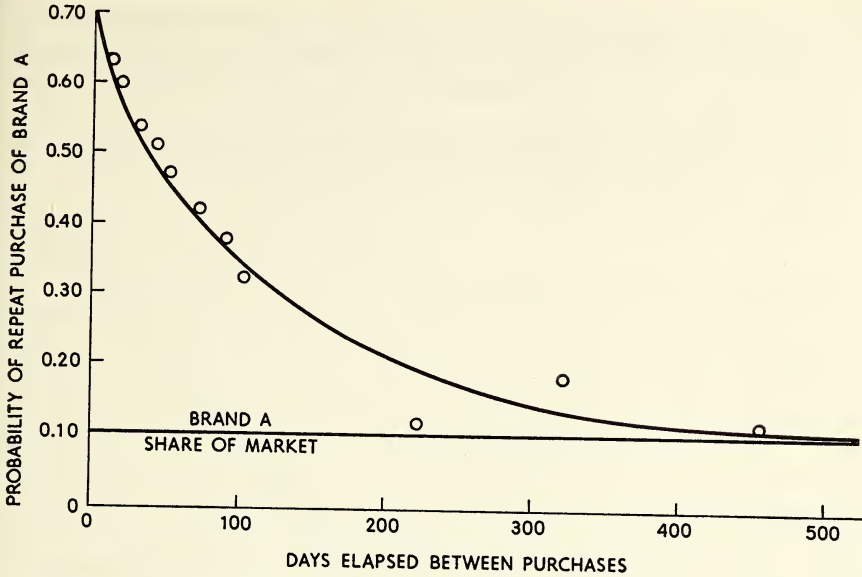


FIGURE 3. The probability of a consumer's buying the same brand on two consecutive purchases of frozen orange juice decreases exponentially with an increase in time between those purchases.

ure 3. Observe also that the probability of repurchasing the same brand at any given time in the future, without regard to the brands chosen in the interim, increases as we go up from $N = 1$ to $N = 3$, $N = 10$, and $N = 50$. Thus, on the average, a consumer who makes his fiftieth pur-

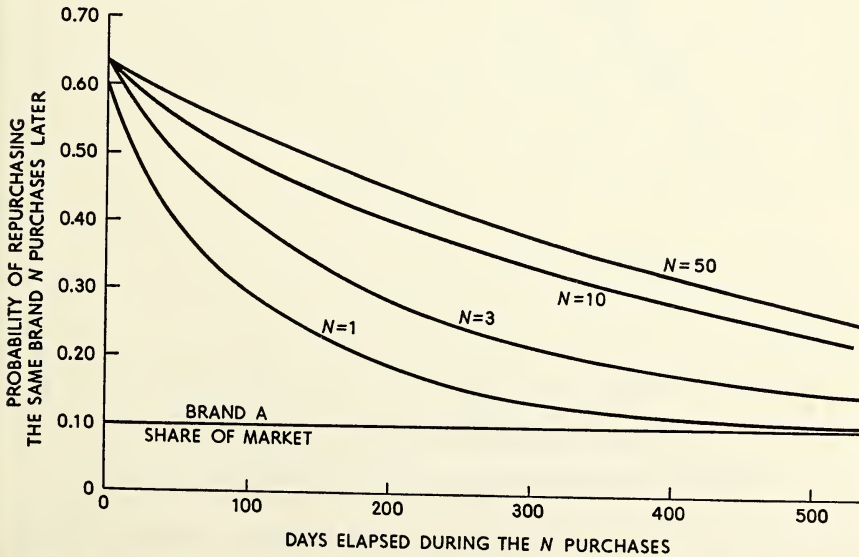


FIGURE 4. Consumers buying frozen orange juice with greatest frequency have the highest probability of continuing to buy the same brand.

chase of frozen orange juice 300 days after some arbitrary purchase of a given brand has a much higher probability of again choosing that brand than does the consumer who makes only 1, 3, or 10 purchases in that interval of time.

Figure 5 illustrates the relationship between the rates of decay of pur-

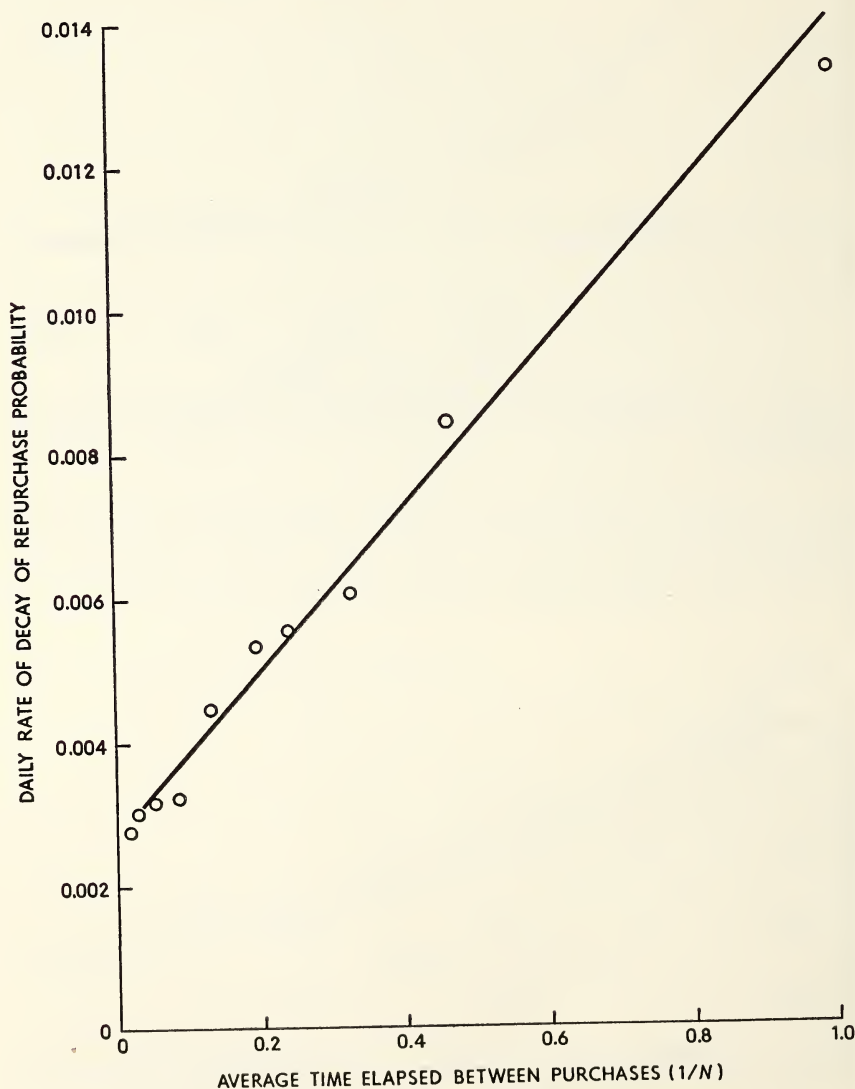


FIGURE 5. Relationship of decay rates to time between purchases.

chase probability associated with the curves in Figure 4 and the average time elapsed between purchases. The rate of decay of $N = 1$ in Figure 4 is 0.01298 per day. The rate of decay of $N = 50$ is 0.00282. Here again

we find a relationship which, because of its simplicity, can after some manipulation be conveniently incorporated into a model forecasting consumer brand choice probabilities. The rate of decay increases linearly with an increase in the average time between purchases. The data points plotted in Figure 5 represent the rates of decay computed for 10 values of N , four of which were illustrated by the curves in Figure 4.

Simulation of Consumer Brand Choice

The brand shifting model outlined in Figure 1 earlier in this paper has been tested by computing the predicted purchase probabilities of consumers on each of approximately 13,000 occasions of purchase of frozen orange juice and comparing aggregates of these predictions with recorded brand purchases. The procedure followed was to first divide the probability space, zero to one, into 76 probability ranges. Then, whenever the computer programed model predicted a certain probability for a given family buying a given brand on a given buying occasion, the results of that purchase were recorded in the computer storage location representing the corresponding probability range. Thus, it was possible to compare within each of the 76 probability cells the average predicted probability of purchasing individual brands with the observed proportion of trials on which the brand was in fact purchased. The predicted and observed probabilities and numbers of purchases were then compared individually and simultaneously for all 76 cells with respect to the binomial and χ^2 distributions that would be expected if the model were perfect. The 76 normal deviates, referred to here by "t," computed for the individual cells with respect to the Snow Crop predictions were approximately normally distributed, 50 lying within 1 standard deviation, 71 lying within 2 standard deviations, and 76 falling within 3 standard deviations. The χ^2 value indicated no significant deviation at the 10 percent level. Similar results were obtained in an analysis of predictions for the Minute Maid brand, 53 "t" values lying within one standard deviation, 70 lying within two standard deviations and all 76 cases falling within three standard deviations.

The above results suggest that the model offers promise for use in describing consumer behavior in probabilistic terms. The model was not tested with respect to individual families, the number of purchases being made by most individual families being considered as providing too small a sample to yield a reasonably powerful test of the predictions of the model. In other words, since rejection is unlikely given a small sample size per family, acceptance does not carry much weight with respect to an evaluation of the model. In the aggregate, the model stood up surprisingly well given the over-all test sample size of approximately 13,000 purchase predictions. (Interestingly enough, when the lower limit L was held at zero in certain tests designed to determine its importance in the model, the "t" values associated with certain low probability cells ranged

from 40 to 65 standard deviations.) Of course, if the sample size were to be increased substantially, significant deviations would have been obtained since the model is not a perfect representation of the brand purchase sequences of consumers.

The predictions of the model were also used to obtain a frequency distribution of consumers throughout the three-year time period according to their probability of buying specific brands of product. Figure 6 pro-

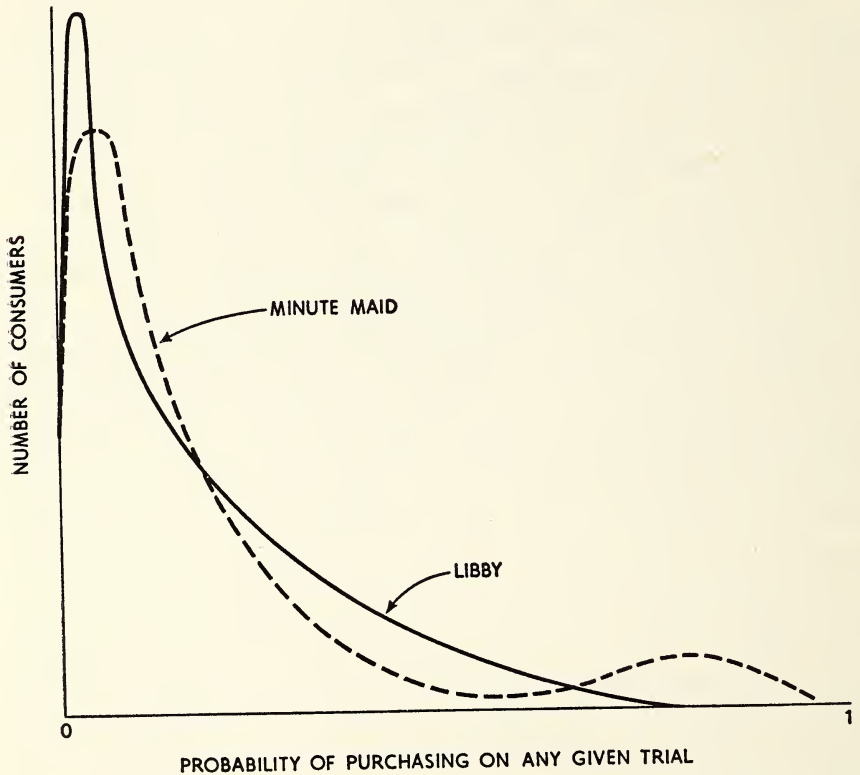


FIGURE 6.

vides a comparison of the profiles (smoothed) for Libby and Minute Maid frozen orange juice. As might be expected, most consumers have a low probability of buying any specific brand. Those consumers who have a high probability of buying one brand must necessarily have a low probability of buying several other brands. Minute Maid was in the enviable position of having a small group of customers with a very high probability of buying the brand. Libby did not have such a following. The fact that Minute Maid developed frozen orange juice and was the first brand available to consumers probably helped the firm develop the group of loyal (or habitual) customers, a sizable portion of which it had been able to retain in the face of growing competition. As the innovator

of frozen orange juice, Minute Maid also developed a pre-eminent market position in terms of retail availability, a factor which undoubtedly helped the firm maintain a sales advantage relative to competition.

ADAPTIVE BEHAVIOR OR SPURIOUS RESULTS?

In a paper titled "Brand Choice as a Probability Process,"⁵ reproduced earlier in this book, Ronald Frank reports that certain results he has observed with respect to repeat purchase probabilities as a function of a brand's run length are similar in appearance to what would be expected with associative learning under conditions of reward. He then notes, in a footnote, that my data also seem to suggest this interpretation, a point on which there is agreement.⁶ The balance of Frank's article is then directed at demonstrating that:

1. Purchase sequence data generated by families for a given brand using a Monte Carlo approach on the assumption that each family's probability of purchasing the brand remained constant throughout the time period produced repeat purchase probabilities as a function of run length which closely approximated in the aggregate the actual observed empirical probabilities.

2. The number of runs observed for *most* families is consistent with what might be expected under the assumption that each family's probability of purchasing any given brand remained constant throughout the time period.

As a result of his success in generating a relationship in (1) above that has the appearance of actual data, Frank states, "These results cast suspicion on the use of a 'learning' model to describe the observations." In view of this statement, which bears directly upon the work I have outlined earlier in this paper, in my thesis, and elsewhere, some defense appears to be in order.

Frank's observations in no way invalidate the findings outlined earlier in this paper. He has shown that it is inappropriate to attribute to learning *all* of the increase in repeat purchase probability associated with increases in run length, an error which has probably been made by more than a few researchers. This is not, however, the approach outlined here or in my thesis. As a matter of fact, the approach used in my thesis could be applied to Frank's coffee data to test whether the probabilities are in fact constant and, if this is not the case, to estimate the appropriate weightings. If consumers were to have a constant probability of brand choice from trial to trial, the most recent purchase positions would not have a greater primary effect on the predicted purchase probabilities than that of any other purchase position—all of the primary effects would be identical except for sampling variations. Similarly, if the probabilities of brand choice were constant from trial to trial, the Purchase and

⁵ Ronald E. Frank, "Brand Choice as a Probability Process," *Journal of Business*, Vol. XXXV (January, 1962), pp. 43-56.

⁶ Alfred A. Kuehn, "A Model for Budgeting Advertising," in Bass, *et al.* (ed.), *Mathematical Models and Methods in Marketing* (Homewood, Ill.: Richard D. Irwin, Inc., 1961).

Rejection Operators in the adaptive brand shifting model outlined at the beginning of this paper would be superimposed on the diagonal (see Figure 1). In other words, the special case considered by Frank can be treated successfully by both of the analytic techniques used in my studies and discussed in this paper. Frank is correct when he states that much of what might appear to be a learning effect on the basis of repeat purchase probabilities as a function of run length is due to the aggregation of consumers having different probabilities (at the start of the run)—this is, however, no problem when one takes into account the effect of all past purchases which have a significant impact upon the consumer's purchase probability, since such an approach does not disregard the information contained in purchases prior to the current run, an important consideration when the run is very short. Since past purchases will, except in highly unusual cases, have decreasing effects (as one goes back in time) upon the consumer's subsequent purchase probability, taking into account all significant past purchases does not generally require the availability of an unduly long record of the consumer's purchase history.

The second point that Frank makes—namely, that most consumers behave as though they had constant purchase probabilities—would appear to represent a misinterpretation of statistical results. Frank sets up his hypothesis, tests it at some level of significance for each of a large number of cases (families), and then interprets the results as though all cases not shown to deviate statistically on an individual basis are consistent with the hypothesis. Actually, the hypothesis was that consumers have a constant probability of purchase, and the results indicated that a larger number of the individual cases tested lay outside the confidence limits than is consistent with the hypothesis, thereby rejecting the hypothesis *in toto!*

To be sure, the hypothesis of constant probability is, in effect, a straw man. It is generally recognized that consumers do change their buying behavior over time. Whether such behavior is called adjustment, adaptation, or learning is unimportant. It should be noted, however, that even though the over-all market for coffee was quite stable in the period studied by Frank, and the sample sizes were limited to 14 months of purchase by each family, the hypothesis was in fact rejected on an over-all basis, the only appropriate way in which to interpret the results of the test. Perhaps, as Frank suggests, some consumers do have constant probabilities of choosing individual brands during certain periods of time. Such a hypothesis cannot be tested, however, unless a procedure independent of the test is available for identifying these consumers and the relevant time periods.

SUMMARY

A model describing brand shifting behavior as a probabilistic process and incorporating the effects of past purchases and time elapsed between pur-

chases has been outlined. A defense of this approach to the study of mechanisms underlying consumer brand choice has also been presented. What has not been discussed is the way in which such merchandising factors as price, advertising, product characteristics, retail availability, and promotions (price off, coupons, merchandise packs, and so on) influence the parameters of the model and the extensions of the model that might be required to incorporate such effects. Some earlier results of research on the influence of these variables have been incorporated into an aggregate "expected value" form of the model presented here. Much work, however, remains to be done.



Multiple Factor Analysis

THE TECHNIQUE OF FACTOR ANALYSIS is used to summarize the information contained in one set of variables in terms of another, smaller, set. The new variables are called "factors" or "principal components"; the analyst attempts to identify them with theoretical constructs which are of interest within the context of his problem. Once such an association has been made, the analysis may be terminated or the new variables may be used as the input for further statistical processes.

Roland Harper's article, "Factor Analysis as a Technique for Examining Complex Data on Foodstuffs," illustrates a natural application of the technique. The problem falls into the general category of product and taste testing: it is much easier to make measurements on the product (in this case cheese) than it is to determine what these measurements really mean. Since many of the observations on the cheeses really measure the same thing, some method for identifying summary factors is desired.

Harper's results show that at least three dimensions are necessary to describe the characteristics of Cheshire cheeses. The reader may ask himself why Harper chooses three, rather than two or four or five dimensions, as being worthy of consideration in view of his factor analytic results. He may also question the extent and significance of the agreement between his results for the two years' data available in the sample. Finally, there is the question of moving from the results of the factor analysis to the specification of tests that can be used to measure the three dimensions of cheese quality in actual quality control work.

In addition to performing the multiple factor analysis upon the Cheshire cheese test data, Harper gives a brief introduction to the technique of factor analysis itself. While written in a more advanced style, his

discussion is basically the same as that presented in "Statistical Analysis of Relations between Variables," elsewhere in this volume.

In contrast to the work of Harper, Dik Warren Twedt's article, "A Multiple Factor Analysis of Advertising Readership," goes beyond the simple interpretation of factor loading coefficients. He uses them as the basis for choosing certain variables for input to a multiple regression analysis. We should note the large number of variables which were included in the factor analysis; since most of the variables are highly correlated with one another, the net contribution of each to the explanation of advertising readership is very difficult to assess. The factor analysis, on the other hand, identifies six distinct new variables which can be used to summarize the original data. Most important of these is the factor "picture and color," the only factor which is significantly associated with readership.

Variables were chosen for use in subsequent multiple regression analyses on the basis of their factor coefficients. Variables that had high coefficients for factors on which the readership coefficient was also fairly high were included in the regressions.

Twedt presents both the original and the rotated factor loadings matrix. The rotations were performed by graphical methods which, while not extensively discussed in the article, imply a considerable amount of subjective judgment on the part of the analyst. Since we can participate vicariously in the naming of the factors (see Table 3), we may ask whether Twedt's results appear reasonable. We can also assess the total contribution to variance made by the six factors to each of the 20 variables. (This figure is called the "communality" of the variable, and is presented in Table 3.) Finally, there might be a question as to whether factor analysis was the appropriate method for analyzing the data in the first place. Relevant here are the number of independent variables, the amount of intercorrelation between them, the linearity of the relations, and the existence (or lack of it) of a set of hypotheses about the effects of each. Since Twedt dealt with many interrelated variables, and without benefit of strong *a priori* notions of effects, it appears that the use of factor analysis was appropriate.

The approach used by Twedt is carried further in the paper by William F. Massy, "Television Ownership in 1950: Results of a Factor Analytic Study." Factor analysis was used to obtain a new set of variables summarizing the information contained in the income and education distributions of a sample of urban places in the United States. The values of the factors were then used as input to multiple regressions and the resulting coefficients utilized in the preparation of an "ownership index" for each cell of the income and education distributions.

Massy's method for applying the factor analytic results to multiple regression inputs was to obtain estimates of the "factor scores" by means of a mathematical procedure based on least squares analysis. (The

factor analytic technique obtains factor coefficients rather than factor scores.) In contrast, Twedt used his factor analysis to select certain variables which appeared to be important; these were used as explanatory variables in his regression equations. The use of a computed factor score is often preferable to selecting and using particular variables because all of the information contained in the factor analysis can then be utilized in the multiple regression.



Factor Analysis as a Technique for Examining Complex Data on Foodstuffs*

ROLAND HARPER†

INTRODUCTION

IN A RECENT ARTICLE IN THIS JOURNAL I DISCUSSED CERTAIN OF THE FUNDAMENTAL problems involved in the subjective appraisal of foodstuffs.⁴ The observations made were not primarily concerned with any one particular food product, although interest in the whole set of problems arose out of a number of years' work on the skills of professional cheese-makers and cheese graders. The question was raised as to how one should examine a set of "complex multi-dimensional relationships" such as may exist between "objective" test data and skilled judgements on a number of cheeses. In the particular investigations to be considered the objective tests were primarily mechanical in nature, but such data might equally well have included information derived from chemical or bacteriological tests. In the previous paper it was suggested that *one* method of dealing with the particular type of problem which has just been posed is the technique of Factor Analysis. A brief outline will be given of the main features of this technique, and its use with food products will be illustrated by these studies of cheeses. It is of interest to add that very recently Baker¹ has used factor analysis to examine the connection between data derived from various forms of routine chemical analysis and data derived from the evaluation of the quality of samples of wine by normal sensory methods. This particular study involved less than twenty samples of wine, whereas almost two hundred cheeses were used in the studies which will be summarised and discussed later.

The main function of the present article is to examine both the potentialities and the limitations of factor analysis as applied to data concerning foodstuffs. No systematic examination along these lines has yet been made. Ostle and Tischer¹⁰ in their otherwise very comprehen-

* Reprinted from *Applied Statistics*, Vol. V, No. 1 (March, 1956), pp. 32-48.

† University of Leeds, Yorkshire, England.

sive review of statistical methods in food research make no reference whatever to factor analysis. Yet it is seven years since the first preliminary study employing data on cheeses was published (Harper and Baron⁵). In what follows an attempt will be made to present the highly technical concepts involved in factor analysis in a form that will be intelligible to those without direct experience of this technique. However, before the aims and functions of factor analysis are reviewed a few observations should be made concerning the origin of its application to cheeses. The investigations that will be used to represent the technique in action were carried out on behalf of the National Institute for Research in Dairying. They owe their origin to the work of Dr. G. W. Scott Blair, in whose department they were carried out. In addition to his pioneer work which led to the development of the various instrumental tests employed, Scott Blair was the first to recognize that factor analysis might lead to a clearer interpretation of the data being accumulated. Specific reference should also be made to Miss Margaret Baron, who was concerned with many different aspects of the field studies. I was concerned mainly with the initial planning and with the final analysis of the data. Five years have now elapsed since I was directly connected with these investigations. Thus it is possible for me to approach the task of evaluating this work with a degree of detachment which would not be possible in the heat of front-line research work.

THE TECHNIQUE OF FACTOR ANALYSIS

This section is concerned with a brief statement in general terms of the aims and the principles underlying factor analysis. It follows along lines already set out in the last chapter of *Foodstuffs, Their Plasticity, Fluidity and Consistency*, edited by Scott Blair.¹¹ Factor analysis should not be confused with factorial design. The latter refers to a systematically planned approach leading to the provision of information about the relative importance of various conditions or processes and the way in which these may interact upon some quantifiable characteristic. There is no danger of the professional statistician confusing these two terms, but they are sufficiently similar for the layman to do so. Many professional statisticians are highly critical of factor analysis, especially on account of its lack of mathematical rigour at certain points. (See Kendall and Babington Smith⁸ for an expert discussion on this theme.) One of the special functions of factor analysis is that of providing a method of sorting out the underlying order in a vast amount of empirical information obtained by different methods in the same general situation. Systematically planned studies require that the main variables of importance are already known. In an exploratory stage this state of affairs has not been reached, and factor analysis may sometimes be employed to isolate the main variables. In this exploratory phase, which would correctly describe the present state of knowledge in connection with the

systematic assessment of foodstuffs, numerous complex issues are involved. In consequence, minor defects in the mathematical rigour of techniques of analysis may be of only secondary importance. Tools should be related to the task in hand, and unrestricted demands for academic precision may be out of place in handling certain complex problems of practical significance. While I admit that factor analysis has its limitations, which will emerge during the discussion of practical examples, our immediate aim is to achieve general understanding.

As already indicated, factor analysis is employed in an attempt to fathom the underlying relationships which may exist between a substantial number of "measures" obtained by testing a number of samples. It is assumed here that several different types of tests are involved. In principle, the samples tested may be almost any class of objects or events. Factor analysis has been elaborated mainly by a number of mathematically minded psychologists; in their own investigations the samples tested have usually been people and the tests employed have been predominantly tests of human performance or mental functioning. Eysenck³ and Vincent¹³ have already discussed certain aspects of factor analysis in this journal, but some recapitulation of the main features is necessary. One convenient method of expressing the relation between two tests is to calculate the correlation coefficient between the "measures" obtained by applying these tests to the same set of specimens. Thus if we denote the two sets of measures by x and y respectively, and make certain assumptions which cannot be elaborated here, the correlation coefficient (r_{xy}) between test X and test Y is given by the expression

$$r_{xy} = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\{\Sigma(x - \bar{x})^2 \cdot \Sigma(y - \bar{y})^2\}^{\frac{1}{2}}} \quad (1)$$

In this expression $\bar{x}$ and $\bar{y}$ represent the averages of the two sets of measures. The numerator in the expression for r_{xy} is simply the sum (over all samples) of the products of the deviations of the x -measure and the y -measure for each sample from the mean values of x and y , respectively, for the whole set of samples. The denominator takes account of the range of variability in the measures. The value of the correlation coefficient calculated in this way will lie between +1 and -1. Plus 1 would denote a perfect linear relationship between the two sets of measures with x and y increasing together. Minus 1 would still indicate a perfect linear relationship, but this time as one set of measures increased the other would decrease in value. Intermediate values of the correlation coefficient denote that the particular measures are imperfectly related. A value of zero would be taken to indicate that no relationship existed. According to the number of specimens tested, the correlation coefficient must exceed a certain value in order to be "significant." This simply means that a certain range of values including zero could arise through the normal workings of "chance" from two sets of measures which are

"really" unrelated. For more detailed information the reader who is not well informed on this topic should consult one of the standard introductory textbooks on statistics.

The procedure outlined for calculating the correlation coefficient between two sets of measures can obviously be applied more or less mechanically to measures derived from a substantial number of tests applied to the same set of specimens. This results in a table of correlation coefficients which is sometimes referred to as a *correlation matrix*. Represented in symbols this would take the form:

<i>Test</i> \ <i>Test</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	...	...
<i>a</i>	—	r_{ba}	r_{ca}	r_{da}	...	...
<i>b</i>	r_{ab}	—	r_{cb}	r_{db}	...	...
<i>c</i>	r_{ac}	r_{bc}	—	r_{dc}	...	...
<i>d</i>	r_{ad}	r_{bd}	r_{cd}	—	...	...
...	...	...	...	...	—	...
...	...	...	...	...	...	—

The diagonal cells have been left blank because they are not normally determined by direct observation. Terms placed symmetrically with respect to the blank diagonal are, of course, identical, but for the sake of completeness they have been represented in a consistent manner; thus:

$$r_{ab} = r_{ba}, \text{ etc.}$$

For certain purposes considerable economy in space is achieved by expressing the original data in this way, although some information is inevitably lost in the process of condensation. Even twenty tests will yield a table containing 190 independent entries. Such a table of numbers may still be too large to grasp and to interpret effectively. Is it possible to produce an even more condensed statement which will represent the underlying order in the data which have been collected? Let us take another purely hypothetical example also expressed in algebraic terms. By means of the computational techniques of factor analysis it is possible to reduce the so-called correlation matrix to a smaller table, referred to as the *factor matrix*. This assumes that there exists a certain degree of order in the data collected. Let us assume that in the present instance this underlying order can be accounted for by three independent, hypothetical "factors." Since we are not at present concerned with the problems of interpretation it is unnecessary to discuss

here the precise meaning to be given to these factors. Under the conditions laid down the factor matrix would take the following form:

Test \ Factor			
	I	II	III
a	a_1	a_2	a_3
b	b_1	b_2	b_3
c	c_1	c_2	c_3
d	d_1	d_2	d_3
...	...	...	...
...	...	...	...

The entries in the table can be interpreted in a number of different ways. For instance, *within certain limits of error*, governed partly by the point at which one stops "extracting further factors," it is possible to reconstruct the original correlation coefficients by a series of equations of the type

$$r_{ab} = a_1b_1 + a_2b_2 + a_3b_3 \quad (2)$$

(assuming that the data do not justify going beyond factor III).

Up to and including instances in which three factors are required it is possible to represent the information contained in the factor matrix by graphical methods or by means of solid models. Beyond this level of complexity some understanding of the geometry of n -dimensions and some acquaintance with matrix algebra seem desirable. In the examples to be considered it will fortunately be unnecessary to go beyond what can be expressed in terms of three dimensions.

Let us consider two further hypothetical examples before turning to real data collected in the field. It is important to bear in mind that the usual methods of factor analysis can lead only to the isolation of what are referred to as "common factors." Several tests are essential for the effective "isolation" of one single common factor. The example taken for graphical representation is one involving two factors (I and II). Attention will be focused upon two tests, a and b, although in practice more than two would be essential to provide the information required to construct these diagrams. Let us consider the manner in which factor analysis allows us to represent the relationship between these tests under two rather different sets of conditions.

Case 1. Representation of the relation between two errorless tests involving two, and only two, common factors.

The entries in the factor matrix corresponding to tests a and b may

be treated as if they were the co-ordinates of the points a and b , as illustrated in the two-dimensional diagram given in Fig. 1. The points a and b have been joined to the centre or origin, and the radiating lines may be referred to as "test vectors." A circumscribing circle of unit radius has also been drawn. This is intended to represent a state of affairs in which any point on the circumference would indicate that the whole of the variation (variance) of that set of measures is completely accounted for by the two common factors. In the present example these are labelled I and II, although it will be evident that the relationship between tests a and b would be preserved even if the reference factors or reference axes were allowed to "rotate" to new positions. Although such rotations of axes are carried out by some investigators they have not been em-

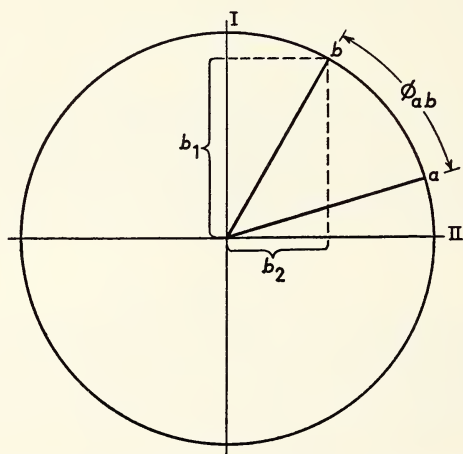


FIGURE 1. Representation of the relation between two errorless tests, a and b , assumed to involve only two common factors. (Reproduced from Fig. 13, section 4.1, of chap. viii of *Foodstuffs, Their Plasticity, Fluidity and Consistency*.)

ployed in the present studies. Hence no further reference to this aspect of factor analysis will be made here. It is customary to denote the length of the respective test vectors (in this case unity) by the symbols h_a and h_b . The square of this length is referred to as the "communality of the test" and indicates the proportion of the test variance accounted for in terms of the common factors which have been isolated. The variance of a set of n measures denoted by the values of x is simply $\sum (x - \bar{x})^2 / (n - 1)$. When more than two common factors are required the full expression for the communality would be:

$$(h_a)^2 = (a_1)^2 + (a_2)^2 + (a_3)^2, \text{ etc.} \quad (3)$$

a_1, a_2, a_3 , etc., being the successive entries in the factor matrix under the headings of factors I, II, III, etc., continued as far as the computational procedures make it necessary. Although in the hypothetical example

given in Fig. 1, h_a and h_b are both unity, this will by no means always be the case. Let us examine this point in relation to Case 2.

Case 2. Representation of the approximate relation between two tests requiring only two common factors.

In this case it is assumed either that the tests involve errors of measurement or that a certain degree of individuality in the tests makes it impossible to account for all the variance in terms of the two hypothetical common factors. This state of affairs is represented in Fig. 2. In this case both h_a and h_b are less than unity. Whereas in Case 1 the correlation be-

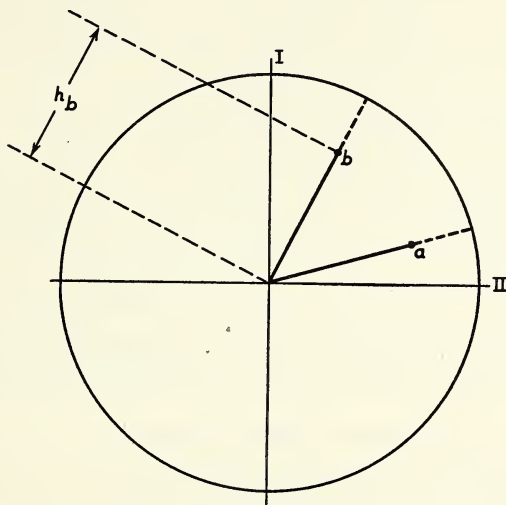


FIGURE 2. Representation of the relation between two tests, a and b , involving error and/or specific variance, and assumed to involve only two common factors. (Reproduced from Fig. 14, section 4.1, of chap. viii of *Foodstuffs, Their Plasticity, Fluidity and Consistency*.)

tween the two tests could be represented solely in terms of the angle between the two test vectors, in Case 2 it is necessary to take into account the lengths of the test vectors. In trigonometrical terms the correlation coefficient—or rather an estimated value based upon the common factors—is given by the equation

$$r_{ab} = h_a h_b \cos \phi_{ab} \quad (4)$$

The same relationship holds good when more than two dimensions are involved. When this happens the angle ϕ_{ab} is an angle out in space. If more than three common factors are necessary this becomes an angle in a generalised space of many dimensions. It will be obvious that the correlation between two tests, ranging as it does between $+1$ and -1 , lends itself to representation in angular form. It could be argued that the techniques of factor analysis enable one to erect a structure within which the relationships between the various tests can be expressed economically in this manner. It is not possible to examine here the geometrical implications of

the various alternative approaches to the general class of problems which have been stated. However, factor analysis is related to a number of the more conventional statistical techniques. A number of important comments on this theme have been made quite incidentally by Cronbach and Gleser.²

It will be evident that certain of the relationships which underlie a number of sets of measures obtained by employing a variety of tests on a set of specimens can be revealed and represented within the framework of factor analysis. No one would wish to claim that this is the only way of revealing such relationships. Nor can it be claimed that factor analysis is capable of dealing with all possible forms of relationship. The relationships which are revealed are not free from such accidental influences as those due to the selection of a *particular* group of tests for probing a situation or a *particular* group of specimens to which these tests are applied. Perhaps it would be better to employ the more general phrase "methods of probing a situation" rather than the more restricted word "tests." Provided that the information can be expressed in quantitative terms, some indication of the many underlying relationships should emerge from the use of factor analysis. Obviously the human senses represent one very important method of probing the world around us, and many examples of this type provide an opportunity for combining objective and subjective data into the same general analysis.

Factor Analysis of Data from Cheshire Cheeses

The origin of the investigations to be discussed has already been indicated. Many of the early studies by Scott Blair and his colleagues employed correlation methods. These included the well-known technique of "partial correlation," whereby it is possible to calculate what the correlation coefficient between two tests would be if the effects of one or more other tests affecting the relationship were held constant. Although this process can be repeated to give high-order partial correlation coefficients, these are extremely difficult to interpret. It appeared that factor analysis might provide a more meaningful overall picture of the relationship between the various tests being employed than was being achieved by the methods of partial correlation. It may be noted that there is a simple relation between the technique of calculating partial correlation coefficients and the technique of extracting factors according to the rules of factor analysis. Although the majority of the earlier cheese studies focused attention primarily upon instrumentally measurable properties, the extension of the test data to include appropriate subjective assessments was simply a further logical development.

Full reports have already been published of the studies of Cheshire cheeses (see Harper and Baron³ for a convenient account of this work). Here it is possible only to present the minimum amount of information essential to subsequent discussion. The field work was carried out in a

commercial cheese factory. The data discussed refer to tests on roughly two hundred cheeses at a time when they would normally have been marketed. The collection of data for the two studies, which were carried out in successive years, was spread over about eight weeks. During this time five cheeses from each day's production were subjected to eight mechanical tests and six types of subjective assessment. The details of the various tests and assessments employed are set out in Table 1.

TABLE 1

<i>Identifying Letter</i>	<i>Description of Test or Assessment</i>	<i>Essential Details</i>	<i>High Values Correspond to:</i>
A	Ball compressor	A measure of surface hardness	Soft, easily indented
-B	Ball compressor	Recovery after removal of load	Small amount of recovery
-D	"Meyer" ball compressor	Measure of "hardening"	No appreciable hardening
E	Borer (rind removed)	Resistance to penetration	Soft, easily penetrated
G	Needle penetrometer (rind removed)	"	"
I	Needle penetrometer (through the rind)	"	"
-K	Spring skewer (through the rind)	"	"
-L	Breaker	Resistance to fracture	Low breaking resistance
M	Firmness (outer surface)	Resistance to pressure of the thumb	Soft
N	Springiness (outer surface)	Recovery on removing pressure	No springiness
O	Firmness (internal sample)	As for M	Soft
P	Springiness (internal sample)	As for N	No springiness
Q	Crumbliness (internal sample)	Breaking down sample with finger and thumb	Lacks crumbliness
R	Overall quality (internal sample)	Includes taste and smell as well as all other characteristics	Poor quality

Negative signs for B, D, K, and L are consistent with those used in previous reports and with the signs used in Table 2.

The instrumental tests (A-L) were concerned with measuring relevant complex physical characteristics. These included surface hardness, resistance to penetration, recovery, hardening as opposed to hardness, and resistance to fracture. Such measurements are relevant for a number of different reasons. Certain of the tests approximate fairly closely to the

manipulative actions of the professional grader when he samples and grades cheeses. The mechanical behaviour of a cheese is relevant here because there is an intimate relation between these characteristics and the balance between moisture content and acidity in the ripening cheese. The subjective assessments (M-R) took the form of simple ratings which involve assigning each cheese to one of five or seven categories intended to represent the range of probable variation under the heading of the particular quality or characteristic being assessed. Although three members of the grading panel were involved here, for the purposes of the present study the individual ratings were simply averaged. In the light of the advanced discussion set forward in the recent contribution to this journal the subjective techniques employed here may appear rather crude (see Harper⁴). However, it must be remembered that the more sophisticated standpoint had not been reached when the present studies were planned and that the introduction of any systematic form of subjective assessment into these investigations was a significant advance upon the design of previous work. The task in hand was to make the most effective use of methods of subjective appraisal which were already to hand rather than to pursue intensive studies of this important facet of the total problem. The assessment of firmness and of springiness was made by pressing the thumb into the surface of each cheese. Such assessments were first made on the outside of the cheese and then upon a cylindrical sample taken from within by means of a grader's sampling iron. Even this action provides the skilled grader with important information about the cheese concerned. The crumbliness of the sample so withdrawn was assessed by breaking down a portion between finger and thumb. Finally, the assessment of overall quality was made. This deliberately took into account taste and smell as well as the characteristics which had previously been assessed.

It is impossible to outline here the various computational stages of factor analysis. Eventually, product moment correlation coefficients were calculated between all pairs of variables, the subjective and the objective data being treated in the same manner. The "centroid method of factor analysis," as described by Thurstone¹² (pp. 149-75), was used to "reduce" the original table of correlation coefficients to a table of factor loadings. Table 2 gives full details of the factor loadings relevant to the two separate studies. It is not necessary to assign meaning to the various factors at this stage. As in the previous hypothetical examples, the factors represent a set of axes at right angles to one another which will be employed to demonstrate the underlying relationships between the various tests. In the present instance there is a considerable degree of correspondence in the two successive years in so far as the first two factors are concerned and a certain measure of correspondence with respect to the third factor. Hence it is possible to represent the salient features of these two separate studies by means of solid models involving three di-

TABLE 2

CENTROID ANALYSES OF SUBJECTIVE AND OBJECTIVE TESTS ON CHESHIRE CHEESES. (FACTOR LOADINGS WITH DECIMAL POINT OMITTED, FOR EXAMPLE +85 REPRESENTS +0.85)

Factor	I		II		III		IV		V		h^2	
	1949	1948	1949	1948	1949	1948	1949	1948	1949	1948	1949	1948
Test	(5 Factors) (4 Factors)											
E	+85	+88	+08	+09	+38	+16	+21	-21	-04	—	916	853
G	+71	+67	+40	+60	+31	+23	+31	-10	-03	—	865	867
Q	+19	+27	+82	+75	+16	+15	-24	-03	+20	—	828	663
O	+85	+84	+22	+25	-23	-10	-09	+25	-25	—	896	846
P	+84	+67	+21	+24	-30	-34	-09	+19	-14	—	866	653
R	+60	+55	+40	+51	-33	-06	-10	-15	+24	—	692	587
-K	+69	+66	-09	-14	+29	+25	-18	+26	-15	—	631	585
I	+75	+76	-34	-31	+28	+12	+06	-11	+06	—	758	704
-D	+65	+44	-51	-51	+17	+11	-21	-20	-04	—	748	507
A	+95	+90	-23	-26	+13	-10	+04	-17	-08	—	976	914
M	+92	+86	-14	-18	-12	+21	+16	+22	+11	—	924	867
N	+90	+74	-15	-25	-18	+12	+10	+12	+14	—	888	642
-B	+82	+62	-44	-68	-20	-28	-12	-21	+05	—	923	961
-L	+64	+39	-24	-15	-38	-40	+15	+09	-17	—	660	346
Variance Accounted for (Percent)	58.2	47.1	12.9	16.7	6.9	4.5	2.7	3.1	2.0	—	82.7	71.4

Data modified from: R. Harper and M. Baron, *British Journal of Applied Physics*, Vol. II (1951), p. 35. M. Baron and R. Harper, *Dairy Industries*, Vol. XV (1950), p. 407; Vol. XVI (1951), p. 45.

mensions. Although such models have been made, mere photographs do not seem to convey effectively the information contained in the model itself. Fig. 3 represents an attempt to convey by means of diagrams what the solid model looks like.

Essentially, the figures given in the first three columns of Table 2 have been used to lay out a set of radiating vectors corresponding to the various tests or assessments. Not only may these vectors vary in

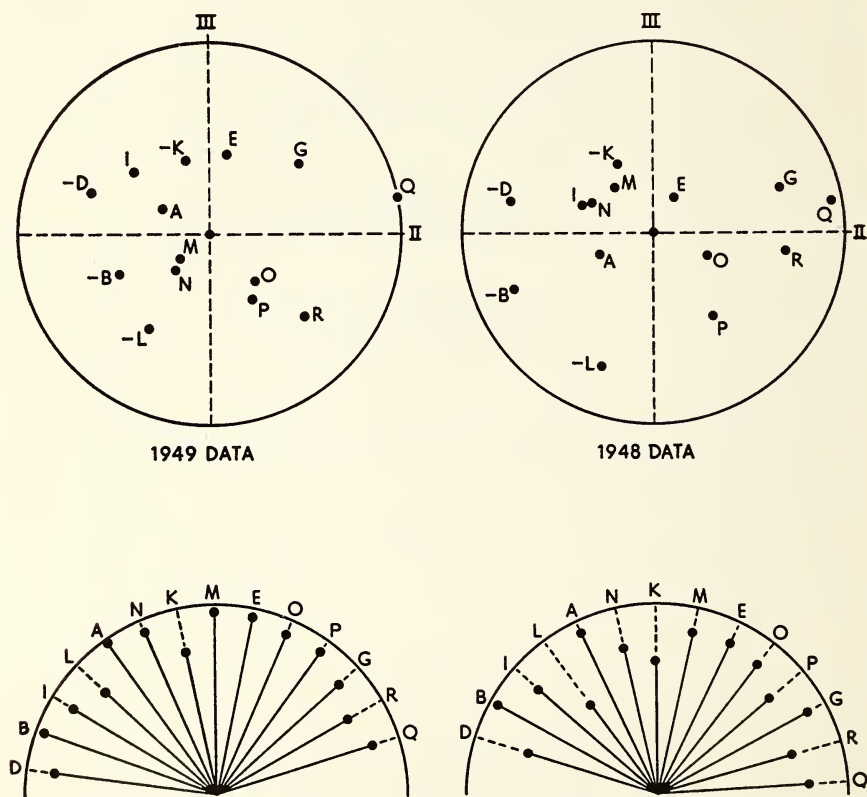


FIGURE 3. Representation of the relations between various tests and assessments of Cheshire cheeses, as indicated by factor analysis. (See text for full explanation.)

angular position but also in length. In the upper half of Fig. 3 the angular relationship is represented by indicating upon a circular plan the points where the various test vectors, if produced, would emerge from the surface of the surrounding unit sphere. It will be recalled that a test vector of unit length implies that the entire variation (variance) of that particular set of measurements can be accounted for in terms of the common factors which have been isolated. The additional information about the lengths of the various vectors is presented in a fanwise manner in the lower half of Fig. 3. Any tests that are represented

by a vector of less than unit length (i.e. the full radius of the sphere) must satisfy one of several conditions. Either the measurements concerned are subject to error or the observed variation in the particular set of measurements must be attributable to factor(s) which are not represented here. The second condition might arise when an errorless test samples characteristics not covered by any other test or assessment in the group. If no "overlapping" test can be found to account for this residual component of variation it would be referred to as "specific variance."

Let us now consider the main points, which are well represented by Fig. 3.

1. Considering that the two studies have been carried out in successive years on a material which is subject to biological variability, the general correspondence is surprisingly good. Certainly there are some minor shifts in the positions and lengths of certain test vectors, and these may require interpretation. However, it might be more correct to consider the data as referring to two different populations rather than to two different samples of the same population of cheeses.

2. These two analyses force us to recognise that at least "three dimensions" are necessary to describe the way in which Cheshire cheeses differ from one another. Of course, this type of multi-dimensional variation is common enough in other fields (cf. hue, brightness, and saturation in the field of colour). In the present instance, partly because of the computational techniques, the particular dimensions represented by the factors I, II, and III are not all of the same importance. As already indicated, there is no need to pin our faith on these accidentally determined factors. It is possible either to "rotate" these axes to a more meaningful set of positions or to pick out specially favoured tests which would effectively provide the complete range of information represented in the three-dimensional picture.

3. The next task is, then, to decide upon the best method of representing these three essential dimensions. There is much to recommend the choice of tests which occupy a unique position in the solid model, e.g. Q, -D, and -L. There is much to recommend the choice of tests involving small errors of measurement. Test A satisfied this condition, for the length of the appropriate test vector is almost unity in both studies. There is also much to recommend the selection of uncorrelated tests so that each will provide information not provided by others. Let us start with test A. In addition to involving a small error component this test comes very close to the position of the "first centroid factor" by means of which a very substantial part of the total variance in all the tests or assessments can be accounted for. Almost at right angles to test A (i.e. largely uncorrelated with it) is test Q, which is also in a unique position in the system as a whole. The identity of these two tests may now be revealed. Test A is an instrumental test of surface hardness which closely simulates the grader's action of "thumbing" a cheese. Test Q is the

subjective assessment of crumbliness, a characteristic for which no adequate instrumental substitute had been invented, at least up to 1950. If we accept these two tests with their traditional basis as adequately representing two of the main dimensions, how can we represent the third? If we insist that the third dimension shall be strictly at right angles to (i.e. independent of) the other two it will be evident that no such test exists. A vector at right angles to tests A and Q would occupy a region in which no actual test vectors are to be found. If any definite statement can be made about the nature of such a hypothetical dimension it is that this must be regarded as a composite variable. All that can be said with any measure of certainty is that this third dimension takes into account the differentiation between variables ($-D$, I , $-K$ and E) and variables ($-L$ and P). Obviously we are moving away from clearly defined interpretation to ambiguous information which seems to offer a whole range of alternative interpretations. The main distinction between these two groups of variables is that between relatively long-time and relatively short-time deformation characteristics. (If this statement is obscure try manipulating "bouncing putty"—a synthetic compound which behaves very differently under different conditions of testing.) However, this is little more than a hypothesis which emerges from the analysis, and which should be submitted to more rigorous methods of testing.

4. A few remarks are necessary about the position of the Overall Quality Variable (R). In the 1949 data the vector representing test R lies closer to the third dimension than any other vector. Unfortunately in this series of studies no attempt was made to secure a separate assessment of flavour. Overall quality is what psychologists would describe as a "global assessment." Wearmouth and Baron,¹⁴ in extending the lines of work which have been outlined to Cheddar cheeses, deliberately included a separate assessment of flavour. Subsequent analysis showed that this showed higher correlation with overall quality than with any other test or assessment. It would be satisfying to suggest that the missing third dimension could best be represented by the assessment of flavour if one were forced to choose an actual test. However, this is a matter for further investigation.

5. The reader may enquire why tests Q , $-D$, and one somewhere near $-L$ have not been proposed as a group for defining and quantifying the most important features of a Cheshire cheese in view of the unique positions involved. Reference to the lower half of Fig. 3 will soon make it clear that the lengths of the vectors representing tests $-L$ and $-D$ fall far short of the desirable unit length, and there is reason to believe on several grounds that these are not the most efficient tests, no matter how advantageous their selection might be from a qualitative point of view.

6. One direct comparison should be made between the Cheshire

cheese studies with those subsequently made on Cheddar cheeses. The majority of tests and assessments were common to both studies. The grading panel was deliberately asked to make separate assessments of both firmness and springiness in the Cheshire studies in spite of the fact that trade experts stated that these two characteristics could not be distinguished. On the other hand, Cheddar cheeses possess a quality known in the trade as "fight-back." This might be described as "the way in which a cheese pushes against your thumb when the pressure has been released." These traditionally recognised distinctions are fully confirmed by the analyses. Table 3 gives comparative information about the same group of four tests; details which are not relevant to the purpose in hand are omitted.

TABLE 3

COMPOSITION OF FIRMNESS AND SPRINGINESS OF CHESHIRE AND CHEDDAR CHEESES
(FACTOR LOADINGS WITH DECIMAL POINTS OMITTED)

<i>Factor</i> <i>Test</i>	<i>Cheshire Cheese</i> ⁶			<i>Cheddar Cheese</i> ¹⁴		
	<i>I</i>	<i>II</i>	<i>III</i>	<i>I</i>	<i>II</i>	<i>III</i>
—	—	—	—	—	—	—
—	—	—	—	—	—	—
M	+92	-14	-12	+73	+19	+32
N	+90	-15	-18	-37	+09	-26
—	—	—	—	—	—	—
—	—	—	—	—	—	—
O	+85	+22	-23	+84	+28	+28
P	+84	+21	-30	+19	+51	-18
—	—	—	—	—	—	—
—	—	—	—	—	—	—

M = Firmness } Assessed by pressing the thumb into the outer surface of the cheese.
 N = Springiness }
 O = Firmness } Assessed by pressing the thumb into a sample taken from within the cheese.
 P = Springiness }

In the studies on Cheshire cheeses firmness and springiness emerge effectively indistinguishable in spite of the fact that these two forms of assessment are not perfectly correlated. In the Cheddar cheese studies the qualitative distinction between firmness and springiness is well substantiated.

Some Critical Comments

An effective appraisal of factor analysis as a technique for examining a set of complex relationships in data on foodstuffs can be carried out under two main headings. The first concerns specifically what has been achieved in these particular studies of cheeses. The second concerns more general questions relating to the application of this technique

to other food products. Factor analysis has a limited range of uses. There may be certain tasks which can best be accomplished by this technique. Perhaps there are others which can be dealt with as effectively, or even more so, by means of some more widely acceptable technique.

It has been argued that the two investigations of Cheshire cheeses have revealed an excellent degree of general agreement in successive years. It has been suggested that the evidence points to the necessity of employing three dimensions in order to describe effectively the way in which one Cheshire cheese differs from another. This refers only to broadly quantitative differences along lines which must be presumed to be common to all Cheshire cheeses. Idiosyncrasies do not emerge from this form of treatment, although they may exist and they may be important. Good reasons have been advanced for using a measure of surface hardness and the subjective assessment of crumbliness to represent the first two dimensions. The difficulties in the way of effectively identifying the third dimension have been discussed, and in spite of a wish to close this issue, the real nature of the third dimension must be left as unsettled. Such would seem to be the main conclusions which have emerged from these time-consuming analyses. However, to have established the multi-dimensional nature of the quality of a Cheshire or a Cheddar cheese is important.

Although no one form of analysis by any means exhausts the many facts lying hidden in the original data, it is not the function of the present article to examine what has been achieved by other methods. The question arises whether conclusions similar to those which have emerged could have been drawn from less time-consuming statistical techniques. Perhaps it is true to say that all types of multivariate analysis are time-consuming, but the necessity for successive approximation by repeating the whole analysis—a regular part of the computational procedure in factor analysis—makes it especially so. Factor analysis could readily be adapted to the electronic computer, and much of the labour involved would thus be eliminated. As a technique for the broad classification of a set of empirical variables without foreknowledge of the nature (structure) of their interdependence, factor analysis seems to offer an approach which cannot be replaced by more conventional statistical techniques. With the same original data it might have been possible to employ the mathematically more rigorous type of analysis into “principal components” (see Hotelling⁷). In this case there is no need to attempt to give meaning to the principal axes which emerge. Since in the present studies emphasis has centred upon tests or assessments rather than factors, perhaps the facts which have emerged might just as well have been expressed in terms of this alternative and more rigorous approach. An analysis along such lines would be essential to test whether this statement is correct or not.

Basic assumptions enter into a technique at a number of different

levels. Assumptions regarding the distribution of the original measures are implicit in the use of the product-moment correlation technique. Further steps in the analysis are based upon the assumption of a number of linear relationships, the effects of which can be added together in a simple manner. To what extent are these and other assumptions sustained by the data being analysed? Some writers believe, largely without proof, that departure from normality in the original measures and minor departures from linear relationships between pairs of variables are of little consequence in factor analysis. However, with this as with any other statistical techniques it is necessary to enquire from time to time whether the basic assumptions implied in the statistical models concerned are satisfied by the data. Unfortunately when a new field of enquiry is being opened up there are so many points at which assumptions have to be made that it is impossible to test whether more than a fraction of them are adequately satisfied. In the present series of investigations it was not possible to examine in detail the distributions of all the measures, nor was it possible to examine whether the relation between all pairs of variables departed significantly from a linear one. In one respect the subjective assessments certainly differed from the instrumental measurements. Since the subjective assessments represented an average of three persons' judgements it is important to note that there are differential restrictions on the possible variance at different parts of the rating scale. Extreme values could be achieved only by perfect consistency between raters, whereas values near the middle of the permitted range could be achieved in a variety of ways. It is impossible to say what influence this might have on the subsequent analysis, but there seemed to be no reason to distinguish between the instrumental tests and the judgements in the final analysis.

Apart from the measure of surface hardness, much remains to be achieved both with regard to the development of additional instrumental tests and with regard to the further standardisation of the methods of making and recording the essential subjective judgements. In view of the lack of finality of such developments, and in view of the many unarticulated assumptions which may be made, it seems reasonable to suggest that the simpler the statistical techniques employed, the better. It used to be fashionable to talk in terms of operational definitions and to say that a particular test itself largely defines what it measures. However, in connection with both the subjective judgements and the instrumental measurements which can be made upon cheeses it is important to stress that an operation is not an isolated event. As Professor Meredith⁹ has indicated operation, operator, and that which is operated upon are inseparably related. Clearly, much remains to be learned about these three aspects of the assessment of foodstuffs. Although some light may have been thrown by factor analysis upon such problems with special reference to cheeses, perhaps more important advances might be made by

fashioning statistical techniques around more clearly defined problems and around the actual (empirically determined) properties of the data being accumulated.

Acknowledgements

I wish to thank Dr. G. W. Scott Blair and Dr. W. G. Wearmouth for helpful discussion, although the opinions expressed are my own. The editors of the *British Journal of Applied Physics* and of *Dairy Industries*, also the North Holland Publishing Company, Amsterdam, Holland, have given their permission to reproduce information which has originally appeared elsewhere.

REFERENCES

1. BAKER, G. A. "Organoleptic Ratings and Analytical Data for Wines Analysed into Orthogonal Factors," *Food Research*, Vol. XIX (1954), p. 575.
2. CRONBACH, L. J., and GLESER, G. C. "Assessing Similarity between Profiles," *Psychological Bulletin*, Vol. L (1953), p. 456.
- ✓3. EYSENCK, H. J. "Uses and Abuses of Factor Analysis," *Applied Statistics*, Vol. I (1952), p. 45.
4. HARPER, R. "Fundamental Problems in the Subjective Appraisal of Foodstuffs," *ibid.*, Vol. IV (1955), p. 145.
5. HARPER, R., and BARON, M. "Factorial Analysis of Rheological Measurements on Cheese," *Nature*, Vol. CLXII (1948), p. 821.
6. HARPER, R., and BARON, M. "The Application of Factor Analysis to Tests on Cheese," *British Journal of Applied Physics*, Vol. II (1951), p. 35.
7. HOTELLING, H. "An Analysis of a Complex of Statistical Variables into Principal Components," *Journal of Educational Psychology*, Vol. XXIV (1933), pp. 417 and 418.
8. KENDALL, M. G., and BABINGTON SMITH, B. "Factor Analysis," *Journal of the Royal Statistical Society, B*, Vol. XII (1950), p. 60.
9. MEREDITH, G. P. "Basic Mental Operations," unpublished address to British Association, Belfast (Section J).
- ✓10. OSTLE, B., and TISCHER, R. G. "Statistical Methods in Food Research," *Advances in Food Research*, Vol. V (1954), p. 161.
11. SCOTT BLAIR, G. W. (ed.) *Foodstuffs, Their Plasticity, Fluidity and Consistency*. Amsterdam: North Holland Publishing Company, 1953.
12. THURSTONE, L. L. *Multiple Factor Analysis*. Chicago: University of Chicago Press, 1947.
13. VINCENT, D. F. "The Origin and Development of Factor Analysis," *Applied Statistics*, Vol. II (1953), p. 107.
14. WEARMOUTH, W. G., and BARON, M. "The Relation between the Mechanical Properties of Cheddar Cheese and Graders' Judgements," *Dairy Industries*, Vol. XVIII (1953), p. 1094.

A Multiple Factor Analysis of Advertising Readership*

.....

DIK WARREN TWEDT†

ONE OF THE MOST WIDELY USED INDICES OF THE ATTENTION VALUE OF published advertisements is the extent to which people read and remember, as determined by readership recognition surveys of the Gallup, Starch, and Advertising Research Foundation type.¹ In these surveys, a representative sample of a publication's circulation (usually from 200 to 400 subjects) is interviewed shortly after publication of the survey issue. Working with a whole copy of the issue (or an abbreviated issue if the original is so large as to cause fatigue during the interview), the interviewer goes through the issue page-by-page, recording the elements which the respondent says he has read. The resulting readership scores are simply percentages of readers who report having read a particular article or advertisement.

The present analysis is based primarily upon the Advertising Research Foundation's Continuing Studies of business magazines (2, 3, 4, 5) for these reasons:

1. They are the most recent studies published by ARF, and have advantages of certain technical refinements such as Lucas' confusion-control (11), and the elimination of respondents who identify more than a critical number of advertisements or articles which have never been published.
2. Business papers which are members of the Audit Bureau of Circulations

* Reprinted from the *Journal of Applied Psychology*, Vol. XXXVI, No. 3 (June, 1952), pp. 207-15.

† Batten, Barton, Durstine, and Osborn, Inc.

¹ Basic data for the present analysis are taken from the Advertising Research Foundation's Continuing Studies of Readership. The ARF, a nonprofit organization sponsored jointly by the American Association of Advertising Agencies and the Association of National Advertisers, has as its purpose the promotion of greater effectiveness in advertising through impartial research. The Foundation makes newspaper, farm paper, transportation advertising, business magazine, and executive management publication readership studies. Since its inception in 1936, the Foundation has published 180 surveys of nine media in 146 markets throughout the United States and Canada (7).

(as are all of the publications represented in the business magazine studies) are required to classify their subscribers by occupation and geographical area (see paragraph 10 of the biannual publisher's statement, available from the Audit Bureau of Circulations or from the publisher).

3. It is reasonable to assume that the population of readers of a business magazine such as *American Builder* is more homogeneous with respect to interest in business advertising, than are readers of general media with respect to consumer advertising.

4. In measuring readership of consumer advertising in general magazines, it is difficult to partial out the cumulative impression of advertising in other media such as radio, television, billboards, etc. This problem is also present in business paper advertising, but to a considerably lesser degree.

Purpose of the Analysis

This analysis has a threefold purpose:

1. To define and measure certain variables in business magazine advertising, and determine the interrelations among these variables, and their relation to readership as measured by the ARF recognition surveys.
2. To determine the factorial structure of the relationships among these variables, so as to make possible a simpler psychological explanation of the obtained variance in readership scores of advertisements.
3. To develop a multiple-regression equation which will predict advertising readership in business papers.

This analysis is thus one of *audience* (or what people do to the advertisements) rather than one of *effect* (what the advertisements do to people). The same general experimental and statistical approach is also applicable to studies of advertising effect, the only stipulation being that an adequate effect criterion must first be available.

The experimental design employed in this study is intended to uncover general principles of advertising which will *increase the probability that prospects will be exposed to a given sales message*. Because of the complex nature of the problem—the many variables which may influence readership both directly and through interaction with other confounding variables—and particularly because of the expense and difficulty of controlled, single-variable experimentation in a practical advertising situation, it is not easy to evaluate the relative importance of variables contributing to variance in readership scores. Comparison of high-scoring advertisements with low-scoring advertisements is helpful, but this does not represent the most powerful statistical technique available. And we do need statistical controls; even where large numbers of observations are available, categorizing the data by such pertinent variables as size and color may reduce the number of cases so greatly that conclusions based upon them are not likely to be stable.

Fortunately there is an exploratory method (multiple factor analysis) which is well suited to the Continuing Studies of readership data. The basic assumption of the factorial method is that there is an underlying order which, when found, will permit us to give a simpler explanation of

phenomena which may seem to be the result of a very large number of variables. The method begins with a table of intercorrelations, or *correlation matrix* (see Table 2). From this matrix we attempt to get simplified explanations or *factors* for the observed correlations (see Table 3).

Procedure

A preliminary analysis was made of the ARF Continuing Study of Business Papers No. 2 (3), on the *American Builder*, a monthly trade magazine edited primarily for building contractors and dealers. At the time of the February, 1950 issue, its circulation was approximately 80,000 (13). The *American Builder* was chosen for this analysis principally because of willing cooperation from the magazine's publisher and research manager.

This magazine averages more than 300 pages to an issue. The survey issue of February, 1950 contained 320 pages, of which only 188 were included in the restapled interviewing copies. The abbreviated survey issue contained 137 advertisements of varying sizes, ranging from $\frac{1}{8}$ page to 4 pages. In advertisements $\frac{1}{4}$ page or larger ($N = 122$), the following readership percentages are available: (1) "Any This Ad," percent who remembered reading or seeing any part of the advertisement; (2) "Headline," percent who remembered reading the principal headline of the advertisement; (3) "Any Copy," percent who remembered reading any of the advertising copy, exclusive of the headlines; (4) "Pictures," percent who remembered seeing the picture indicated. For advertisements smaller than $\frac{1}{4}$ page, only one readership percentage, "Any This Ad," is given. For all advertisements $\frac{1}{4}$ page or larger, "Any This Ad" readership percentages correlated .98 with "Pictures," .91 with "Any Copy," and .90 with readership of "Headlines." The more inclusive category, "Any This Ad," was chosen as the criterion measure of readership.

Against this criterion, product-moment r 's were computed for 34 advertising variables (see Table 1). Mechanical variables are listed as items 1 through 15 in Table 1, and content variables are listed as items 16 through 34. Detailed definitions of each variable have been deposited with ADI, from which microfilmed copies are available at nominal cost.²

The correlations of .00 and .01 between readership and Flesch readability indices (10, 16) were not statistical artifacts due to restriction in range of Flesch scores, but they may be a function of high specialization of interest by technical audiences.

Of the 34 variables which were correlated with the readership criterion, 19 variables were selected on the basis of significant correlation with the criterion. In Table 1, variables 6, 23, 25, 32, and 33 were not in-

²Detailed definitions have been deposited with the American Documentation Institute. Order Document 3417 from American Documentation Institute, 1719 N. St., N.W., Washington 6, D.C.

cluded in the correlation matrix because they were not independent of other variables which were included. Variable 28 was not included because of low reliability of judges. Means and standard deviations are included in Table 1 to provide the reader with some knowledge of the distributions from which the correlations were obtained. The units in which the means and σ 's are expressed are described in the ADI section of this paper. Product-moment intercorrelations were computed for these

TABLE 1
CORRELATIONS WITH READERSHIP, MEANS, AND STANDARD DEVIATIONS OF 34 ADVERTISING VARIABLES

<i>Variable</i>	<i>r</i>	<i>M</i>	<i>σ</i>
K. Criterion (Percent Readership)*.....		26.7	16.6
<i>Mechanical Variables</i>			
1. Number of pages or size of advertisement.....	0.62†	0.6	0.6
2. Width-height ratio of advertisement.....	0.15	0.6	0.3
3. Number of colors.....	0.37†	2.3	0.5
4. Number of separate illustrations.....	0.28†	2.7	3.9
5. Square inches of illustration.....	0.67†	12.4	17.5
6. Proportion of illustration.....	0.52	24.3	16.4
7. Number of type styles.....	0.06	4.9	1.6
8. Number of type sizes.....	0.28†	5.6	1.6
9. Point size of largest type.....	0.49†	38.9	23.1
10. Point size of headlines (weighted average).....	0.43†	20.7	12.1
11. Largest type: product identification.....	0.39†	27.7	19.8
12. Point size of main body copy.....	0.24†	7.9	1.2
13. Pica width of copy measure (weighted average).....	0.35†	14.6	5.5
14. Number of copy blocks.....	0.30†	6.7	4.0
15. Layout deviation ($\pm$) from 90 degrees.....	-0.13‡	12.3	6.7
<i>Content Variables</i>			
16. Flesch readability scores.....	0.00	46.9	10.7
17. Flesch abstraction level scores.....	0.01	27.3	7.1
18. Number of words in advertisement.....	0.31†	169.3	162.0
19. Number of words in headlines.....	0.10	11.3	13.8
20. Number of product identifications.....	0.40†	6.8	5.3
21. Number of product facts.....	0.19†	16.5	11.5
22. Number of product benefits.....	0.29†	11.6	8.9
23. Number of pictorial benefits.....	0.18	1.6	2.2
24. Number of benefits in headlines.....	-0.05	1.7	1.8
25. Number of benefits in body copy.....	0.28	8.6	6.7
26. Number of pictures of product not in use.....	-0.33†	2.3	1.7
27. Directions for getting more details.....	-0.09	2.0	0.7
28. News value ratings.....	0.29	1.6	0.7
29. Readership of surround.....	0.32†§	30.5	12.6
30. Number of similar ads in issue.....	-0.04	18.3	16.6
31. Previous schedule: 1/50 + 1949.....	0.47†	5.0	5.8
32. Previous schedule: 1/50 + 1949-48.....	0.45	9.2	13.3
33. Previous schedule: 1/50 + 1949-48-47.....	0.45	13.1	17.7
34. Brad-Vern totals.....	0.23†	38.3	105.0

* r 's based on advertisements in Continuing Studies of Business Papers No. 2, *American Builder*, issue of February, 1950. N = 137 unless otherwise indicated.

† Significant r 's which are included in the 20 $\times$ 20 correlation matrix.

‡ Based on 57 full-page advertisements.

§ Based on 34 advertisements.

TABLE 2
CORRELATION MATRIX OF PRODUCT-MOMENT r 's FOR 20 ADVERTISING VARIABLES*

Variable	K	2	3	4	5	8	9	10	11	12	13	14	18	20	21	22	26	29	31
K. Readership																			
2. Size of ad.																			
3. Number of colors.																			
4. Number illustrations.																			
5. Square inches of illustration.																			
8. Number type sizes.																			
9. Largest type.																			
10. Headline size.																			
11. Largest product identifications.																			
12. Body type size.																			
13. Pica width.																			
14. Number copy blocks.																			
18. Number words.																			
20. Number product identifications.																			
21. Number product facts.																			
22. Number product benefits.																			
26. Pictures of use.																			
29. Surround.																			
31. Previous schedule.																			
34. Brad-Vern schedule.																			

* Decimals omitted.

19 variables plus the readership criterion, and incorporated in a 20×20 correlation matrix (Table 2). In general, the variables are positively correlated.

The correlation matrix was factor analyzed with Thurstone's complete centroid method (15, Ch. VIII). The resulting centroid matrix is shown in Table 3. Extraction was stopped with the sixth factor, since the product of the two highest loadings in factor VI ($.27 \times .30 = .08$) is only equal to the standard error of the original r between these two variables.

In order to give psychological meaning to these factor loadings, the arbitrary reference frame obtained by the centroid method was rotated by the graphical method. The resulting factors are orthogonal. Criteria of positive manifold and simple structure were observed wherever possible. Since most of the correlation coefficients are positive, it might be expected that a positive manifold could be obtained, and this actually was achieved with only a few exceptions.

Interpretation of Factors

In Table 3, boldface figures indicate the factor on which each test variable has its highest loading (a factor loading is the correlation between that test and that factor). Coefficients of determination (the squared factor loadings) give the percentage of variance of a given measurement which may be predicted by a particular factor. For example, the Readership variable has a loading of .64 on factor *PC*; thus $.64^2 = .41$, or 41 percent of the variance in readership scores may be predicted from this single factor. Note that only two of the factors (*Pictorial-Color* and *Size*) have major loadings on Readership. Factor loadings below .20 are not usually considered significant; loadings between .30 and .40 may be important; if the projections are .40 or above, the loadings are considered significant.

Factor PC has high positive loadings on Readership (.64), Square inches of illustration (.51), Number of pictures showing the product in use (.51), Number of colors (.49), and Previous schedule of advertising (.42). The best measures of this factor are those involving *Pictorial* and *Color* aspects of advertisements, hence the factor designation *PC*.

Factor S has high loadings on Ad size (.69), Number of product benefits (.52), Square inches of illustration (.48), Number of words (.46), Previous schedule of advertising (.46), and Largest type size (.45). Readership loading on factor *S* is .35, or 12 percent of the variance in readership scores is attributable to this factor, which seems to involve *Size* of advertisement.

Factor T has high loadings for Largest type size (.62), Readership of surround (.61), Largest type used for product identification (.48), Number of type sizes (.47), and Point size of main body copy (.46). In general, this factor seems to be associated with *Typographic* size and

TABLE 3

FACTOR LOADINGS AND COMMUNALITIES OF 20 ADVERTISING VARIABLES*

Variable	Centroid Loadings							Rotated Loadings						
	I	II	III	IV	V	VI	h^2_c	PC	S	T	I_n	F	A	h^2_r
K. Readership.....	.74	-18	-21	-18	08	-12	6773	64	35	28	18	16	09	6766
5. Square inches of illustration.....	.08	-13	-10	-13	14	-30	6158	51	48	25	06	23	04	6111
26. Pictures of use.....	.38	37	11	-50	05	-10	5559	51	23	-18	09	10	-44	5571
3. Number of colors.....	.33	-25	-17	-31	-15	13	3358	49	-07	23	11	-15	01	3326
1. Size of ad.....	.87	07	18	30	27	-10	9827	18	69	26	45	45	04	9827
21. Number product facts.....	.35	48	15	-17	-21	-07	4533	20	37	-24	28	-12	-35	4498
34. Brad-Vern schedules.....	.26	19	-39	19	02	-27	3652	21	37	-25	-01	07	34	3641
9. Largest type.....	.71	-36	29	18	00	-06	7538	15	45	62	36	12	03	7543
11. Largest product identifications.....	.51	-33	16	08	-14	-09	4287	18	34	48	21	-06	04	4277
8. Number type sizes.....	.55	-18	46	11	-28	07	6419	01	37	47	47	-16	-19	6405
12. Body type size.....	.26	-47	-09	-04	-02	04	3002	12	04	46	09	00	25	2982
10. Headline size.....	.55	-27	-10	18	06	04	4230	26	24	34	30	14	27	4233
18. Number words.....	.64	38	15	30	-18	17	7278	05	46	-10	71	00	-03	7291
14. Number copy blocks.....	.62	43	21	-08	-04	27	6943	25	27	-13	65	10	-33	6937
20. Number product identifications.....	.66	37	-11	14	04	25	6683	31	29	-20	64	19	06	6695
22. Number product benefits.....	.60	59	13	14	-22	04	6965	14	52	-22	57	-06	-17	6958
29. Surround.....	.24	-54	18	20	75	16	10097	-03	-13	61	10	76	17	10064
31. Previous schedule.....	.55	16	-53	25	04	-19	7092	42	46	-21	19	14	47	7087
13. Pica width.....	.34	-44	-33	08	-07	11	4415	35	-01	34	14	05	43	4452
4. Number illustrations.....	.46	27	15	-34	18	07	4599	36	17	-05	27	25	-40	4564

* Decimals omitted. Boldface indicates factor on which each variable has its highest loading.

variety. Its Readership loading is .28, accounting for 8 percent of readership variance.

Factor In has high loadings for Number of words (.71), Number of copy blocks (.65), Number of product identifications (.64), Number of product benefits (.57), Number of type sizes (.47), and Ad size (.45). This factor appears to be one of *Information*, and its loading on Readership is .18, accounting for only 3 percent of readership variance.

Factor F has only two significant loadings: Readership of surround (.76) and Ad size (.45). The factor designation F is for *Field*—the influence of the surrounding field or background against which the advertisement is seen. Another 3 percent of readership variance is accounted for by this factor, which has a Readership loading of .16.

Factor A has significant loadings for Previous schedule (.47), Number of pictures of product in use (−.44), Pica width of copy measure (.43), and Number of illustrations (−.40). This factor is difficult to interpret, but tentatively it is called A, for *Advertising* schedule previously run. It accounts for less than 1 percent of readership variance; the criterion loading is .09.

An important conclusion is that collectively these six factors account for two-thirds ($b^2 = .6766$) of the observed variance in readership scores of advertisements appearing in the February, 1950 issue of *American Builder*. PC alone accounts for 41 percent of the variance in readership scores; PC and S together account for 53 percent of the variance.

Prediction of Readership from Multiple Regression Equations

On the basis of the factor analysis, certain variables were chosen which seemed to be factorially purest, and which also offered most promise for prediction of advertising readership. Several combinations of these variables were tried in multiple regression equations, and the following set of three (see Table 4) was selected as providing maximum prediction

TABLE 4
CORRELATION MATRIX OF THREE ADVERTISING VARIABLES

Variable	Size	Colors	Square Inches of Illustration
Size of advertisement.....		−0.07	0.71
Number of colors.....	−0.07		0.21
Square inches of illustration.....	0.71	0.21	

with minimum trouble of measurement.

$R_{1.234} = .77$ (where 1 = Predicted readership; 2 = Size of advertisement; 3 = Number of colors; 4 = Square inches of illustration). Correction for bias gives a shrunken R of .76. When nine variables (numbers

1, 3, 11, 20, 21, 22, 29, 31, 34) were incorporated in a regression equation,³ $R = .79$. The gain of .03 is obviously not worth the time involved in making these additional measurements.

The best comparison of each variable's contribution to the variance in readership is found in column (4) of Table 5, where each beta weight

TABLE 5
CORRELATIONS WITH READERSHIP, β WEIGHTS, βr CROSS PRODUCTS, AND
REGRESSION COEFFICIENTS OF THREE ADVERTISING VARIABLES

(1) <i>Variable</i>	(2) r_{1k}	(3) β	(4) r	(5) b_{1k}
Size of advertisement.....	0.62	0.441	0.273	8.293
Number of colors.....	0.37	0.341	0.126	3.869
Square inches of illustration.....	0.67	0.285	0.191	0.181

is multiplied by the corresponding raw r . The regression coefficients, or optimal weights by which each variable must be multiplied to obtain a maximum multiple R , are given in column (5).

The regression formula computed from the *American Builder* data is:

$$X' = 10.456 + 8.293 (\text{Size of ad in pages}) + 3.869 (\text{Number of colors}) + .181 (\text{Square inches of illustration}),$$

where X' = predicted readership, and 10.456 is a correction for point of origin. For correlational purposes, of course, this constant may be eliminated from the computations. It is also obvious that prediction of readership by this formula establishes *relative* differences rather than absolute readership scores, which are dependent upon the general readership level of a particular magazine.

In order to minimize the possibility of computational error in the calculation of R and the appropriate regression coefficients, a product-moment correlation was computed between actual readership scores of the 137 advertisements in the *American Builder* study, and predicted readership scores of these advertisements, based upon the regression weights given in Table 5, column (5). The correlation coefficient was .76—agreeing with the shrunken R of .76.

The critical point of the study is now at hand: the factorial approach proved fruitful with the *American Builder* data, but what is the strength of the relationship between the mechanical variables of size, color, and amount of illustration and readership of advertising in *other* business magazines? Table 6 shows the product-moment correlation coefficients between actual readership scores of advertisements in other

³ Thorndike's (14, p. 340) adaptation of the Kelley-Salisbury iterative solution for R greatly facilitated these computations.

TABLE 6

CORRELATIONS BETWEEN READERSHIP SCORES OF ADVERTISEMENTS
IN ARF STUDIES, AND READERSHIP SCORES PREDICTED FROM THE
REGRESSION FORMULA

<i>Magazine Surveyed</i>	<i>r</i>	<i>Number of Advertisements</i>
<i>Automotive Industries</i>	0.58	131
<i>American Builder</i>	0.76	137
<i>American Machinist</i>	0.63	161
<i>Chemical Engineering</i>	0.64	133
<i>Business Week</i>	0.80	101
<i>Successful Farming</i> :		
Men readers	0.77	217
Women readers	0.73	217

ARF studies (1, 2, 3, 4, 5, 6) and readership scores predicted from the regression formula.

The mean r for the four business magazine studies is .66 (obtained by Fisher's z transformation). When *Business Week* (6) and *Successful Farming* (1) are included with the business magazine studies, the resulting mean r is .71.

Discussion of Results

The three possible sources of variance in readership of advertisements are: (1) differences in the attention-getting power of the advertisements; (2) differences in respondents' interests and purchasing readiness; and (3) chance errors in measurement.⁴ The present study is concerned only with readership variance attributable to differences in the advertisements, whether these differences are *mechanical* (size, color, illustration, etc.) or differences in *content* (number of facts, benefits, etc.).

When this analysis was begun, a possible outcome was that only a small part of the differences in readership might be accounted for by mechanical variables. A recent evaluation of the importance of *content* as against *mechanical* variables has been given by James D. Woolf (17, p. 43), formerly vice-president of the J. Walter Thompson advertising agency, who stated:

"It is my conviction that it isn't the size of the space that puts PULL into an advertisement. At least size is not the most vital consideration. Dr. Samuel Johnson said two centuries ago that 'the soul of the advertisement is the size of the promise.' In other words, the size of the promised benefits.

⁴This trichotomy is somewhat oversimplified; the possibility also exists that for two given advertisements, A might have greater immediate attention value than B, and yet B might be remembered more readily than A several days after S's original exposure to A and B. Thus A might be said to have greater *attention value*, but B greater *memorability*. In the present study, these two variables (if they actually do exist independently) are confounded and their effects cannot be measured separately.

"I have no doubt that the huge units of space tend to achieve certain desirable results for the advertiser, they stimulate the sales force, they impress the trade, and very likely they enhance the importance of the product in the eyes of the reader. But they do not shape public opinion for a product when and if Dr. Johnson's 'soul of the advertisement' is not in the copy. Huge size and red ink and thunderous pitch and clamor are not substitutes for promised benefits. . . ."

The italics are Woolf's. This position is clearly not supported by the findings of the present study. It should be remembered however, that Woolf may refer to consumer advertising, and that he may exclude industrial advertising from consideration (although he does not make this distinction explicit in the article from which this quotation is taken) and secondly, his undefined "PULL" may refer to *effect* of the advertisement (which this study clearly does *not* attempt to predict) rather than to the size of its audience.

Lucas and Britt (12, p. 289) also stress the importance of content: ". . . the primary element in the success of all advertising copy is its *content* or substance. Most of the other factors are merely devices for making the subject matter more visible, more palatable and easier to comprehend."

In an earlier experiment which was designed to measure the influence of mechanical variables on readership of newspaper advertisements, Ferguson (9) concluded, "Contrary to popular and scientific belief it was found that there was no relationship between the size of an advertisement and its attention value." Ferguson's data were based on readership of a small daily newspaper, and again there may be real differences between readership of industrial advertising in business magazines and consumer advertising in small daily newspapers.

Although the present conclusions as to the importance of the mechanical variables of size, color, and illustration are based primarily upon industrial advertising in business magazines, it is suggestive that the highest relationship between readership scores, and readership as predicted by the regression formula, was for *Business Week*, an executive management publication which is somewhere in between the business magazine edited for a particular industry or occupation, and the general magazine with almost universal appeal. Unless this r of .80 (Table 6) represents only a vagary of sampling, it is reasonable to assume that the regression weights given in Table 5, column (5) may prove useful in predicting the relative readership of advertisements in general magazines. The *Successful Farming* study, with r 's in the .70's (Table 6), also supports this assumption.

SUMMARY

Thirty-four advertising variables were defined, measured, and correlated with readership scores for 137 advertisements in the February, 1950 issue of

the *American Builder*, a business magazine published primarily for building contractors. Criterion scores were obtained from the Advertising Research Foundation's Continuing Studies of business magazine readership.

Of these 34 variables, 19 were selected as most significantly correlated with the criterion. Product-moment intercorrelations were computed for these 19 variables plus the readership scores, and the resulting 20×20 correlation matrix was factor analyzed.

Six factors were found to be sufficient to account for the intercorrelations. Of these six factors, only two, *PC* (Pictorial-Color) and *S* (Size) have major loadings on Readership. The other factors are *T* (Typographic size and variety), *In* (Informational), *F* (Field factor, or the influence of the surrounding field of the advertisement), and *A* (Advertising schedule previously run). Collectively, these six factors account for two-thirds of the observed variance in readership scores of the advertisements. The *PC* and *S* factors alone account for 53 per cent of the variance in readership.

On the basis of the factor analysis, certain variables were chosen which seemed factorially purest, and a multiple regression equation was developed to predict readership of advertisements in other business magazines. A multiple *R* of .77 was obtained between readership and the following group of variables: size of advertisement, number of colors, and square inches of illustration.

The regression equation was employed to predict readership of advertisements in six other Advertising Research Foundation studies. Predicted readership scores were correlated with actual readership scores, and these validity coefficients ranged from .58 to .80, with an average *r* of .71.

REFERENCES

1. ADVERTISING RESEARCH FOUNDATION. *Continuing Study of Farm Publications*: No. 3, *Successful Farming*, issue of May, 1947. New York: ARF.
2. ADVERTISING RESEARCH FOUNDATION. *Continuing Study of Business Papers*: No. 1, *Automotive Industries*, issue of October 15, 1948. New York: ARF.
3. ADVERTISING RESEARCH FOUNDATION. *Continuing Study of Business Papers*: No. 2, *American Builder*, issue of February, 1950. New York: ARF.
4. ADVERTISING RESEARCH FOUNDATION. *Continuing Study of Business Papers*: No. 3, *American Machinist*, issue of March 6, 1950. New York: ARF.
5. ADVERTISING RESEARCH FOUNDATION. *Continuing Study of Business Papers*: No. 4, *Chemical Engineering*, issue of March, 1950. New York: ARF.
6. ADVERTISING RESEARCH FOUNDATION. *Study of Executive Management Publications*: No. 1, *Business Week*, issue of April 22, 1950.
7. ADVERTISING RESEARCH FOUNDATION. "The Advertising Research Foundation: What It Is—What It Does." New York: ARF, 1951.
8. BRAD-VERN'S REPORTS. New York: Printers' Ink Publishing Company, 1950.
9. FERGUSON, L. W. "The Importance of the Mechanical Features of an Advertisement," *Journal of Applied Psychology*, Vol. XIX (1935), pp. 521-26.
10. FLESCH, R. "Measuring the Level of Abstraction," *Journal of Applied Psychology*, Vol. XXXIV (1950), pp. 384-90.
11. LUCAS, D. B. "A Rigid Technique for Measuring the Impression Values of Specific Magazine Advertisements," *Journal of Applied Psychology*, Vol. XXIV (1940), pp. 778-90.

12. LUCAS, D. B., and BRITT, S. H. *Advertising Psychology and Research*. New York: McGraw-Hill Book Co., Inc., 1950.
13. STANDARD RATE & DATA SERVICE, Business Publication Section, Vol. XXXII (1950), p. 116. Chicago.
14. THORNDIKE, R. L. *Personnel Selection*. New York: John Wiley & Sons, Inc., 1949.
15. THURSTONE, L. L. *Multiple-Factor Analysis*. Chicago: University of Chicago Press, 1947.
16. TWEDT, D. W. "A Table for Use with Flesch's Level of Abstraction Readability Formula," *Journal of Applied Psychology*, Vol. XXXV (1951), pp. 157-59.
17. WOOLF, J. D. "It Isn't Size That Puts Pull in Advertising," *Advertising Age*, April 30, 1951, p. 43.

Television Ownership in 1950: Results of a Factor Analytic Study*

.....

WILLIAM F. MASSY†

WHAT KINDS OF PEOPLE ARE THE FIRST TO BUY PARTICULAR CONSUMERS' durable products? For the case of television, this paper attempts to answer the question by factor analyzing cross-sectional data on the distributions of independent characteristics of the population, and relating the resulting factors to percentage television ownership. It reports results from an early phase of a long-term study of the growth of television ownership in the late nineteen forties and early fifties.

THE NEED FOR SENSITIVE ESTIMATION PROCEDURES

The aggregate fraction of the population in any geographic area who own television receivers is made up of the percentage of ownership of population subgroups, weighted by the group's relative frequency in the population. Study of the growth of the television market makes sensitive measures of these subgroup percentage ownership figures desirable. (Yet only area-wide aggregate data on percentage ownership are available.) If income and education are relevant explanatory variables, for instance, we would like to be able to assess the propensity to own television receivers for families with particular levels of income and educational achievement. Information on the TV ownership of families falling into each of a number of fairly fine divisions of the income and education distributions would be desirable.

The use of cross-sectional data for an analysis of market growth requires that demand characteristics for areas at different stages of market

* The author acknowledges generous grants of computer time made by the M.I.T. Computation Center, Cambridge, Massachusetts (M.I.T.-IBM 7090 Computer), and the Sloan Research Fund of M.I.T.'s School of Industrial Management, which supports the S.I.M.'s IBM 1620 computer facility.

† Stanford University.

development be compared.¹ This, in turn, implies that statistically stable measurements of demand characteristics be obtainable from a relatively small number of cross-sectional observations: if a sample is divided into classes on the basis of stage in market development, the number of observations in each class is likely to be small.

The foregoing considerations suggest that the traditional aggregate explanatory variables, such as median income, should be supplemented with additional information. One source of information is the *distribution* of such variables in the population.² The approach utilized in this paper makes use of detailed data on the proportion of households in each cell of the income and education distributions for geographic areas, as reported in the U.S. census of population. The contribution of families in each cell to the area's aggregate television ownership will be estimated. The procedure is extremely flexible: distribution data for additional variables can easily be incorporated into the framework when desired.

EARLY TELEVISION DEMAND

Television was first seriously introduced to the United States market in 1946, following development efforts in the thirties and the tremendous technical advances in electronics achieved during the war years. It was an almost immediate success: in the four years leading up to the 1950 census of population the number of homes equipped to receive television broadcasts grew from almost zero to over four million.

The speed of diffusion of this innovation, as well as its intrinsic importance, led a number of analysts to measure television's demand characteristics during its early period. The most extensive study reported to date has been the one performed by Dernburg.³ He used regression and analysis of variance methods to ascertain the relationship between percentage ownership of television and: (1) median educational achievement, median personal income received by families and unrelated individuals in 1949, the relative quartile deviation of income, and a number of demographic variables; and (2) the age of the oldest station received in the area (a measure of time since market penetration began), and the

¹The need for measuring differences in the structure of demand, for areas in various stages of the life cycle of an innovation, is discussed extensively in William F. Massy, "Innovation and Market Penetration, A Study in the Analysis of New Product Demand" (unpublished Ph.D. dissertation in economics, Massachusetts Institute of Technology, September, 1960).

²Micro data on ownership or purchases, by individual family, would of course be ideal. Then ownership could be compared with the socioeconomic characteristics of the particular consumer unit in question. Unfortunately, historical data on problems of interest are usually available only in the aggregated form discussed in this paper.

³Thomas F. Dernburg, "Consumer Response to Innovation: Television," in Dernburg, Rosett, and Watts, *Studies in Household Economic Behavior*, Yale Studies in Economics, Vol. IX (New Haven, Conn.: Yale University Press, 1958).

number of television signals available during the years preceding 1950 (a measure of program variety). His unit of observation was the census tract, which contains from 500 to 10,000 or more dwelling units. Data on approximately 3,000 tracts were obtained from the 1950 census of population.

Dernburg found that, as of 1950, the length of time and intensity of television coverage, together with income, explained about 65 percent of the variance of percentage television ownership among census tracts. Increasing percentage ownership was associated with increasing median income up to levels of between \$6,500 and \$7,500; but TV appeared to be an inferior good above these levels. The same kind of result held for education: TV ownership tended to be largest for tracts having median education levels of from 10.0 to 11.9 years (that is, the achievement of persons who had not quite finished high school), and to decline both above and below these values.

Dernburg also showed that except for tracts with very low median incomes an increase in the dispersion of income, as measured by its relative quartile deviation, was associated with a reduction of percentage ownership. This result is compatible with the hypothesis that television is an inferior good for high income groups, since for most areas an increase in income dispersion (with median income held constant) is an indication of larger frequencies in the high income ranges.

There was a positive association between set ownership and family size, but the probability of set ownership declined as the proportion of children or aged persons in the family increased. Television was inversely related to the relative size of the nonwhite population. It was positively associated with the relative number of housewives in the tract.

DESCRIPTION OF THE SAMPLE

The study reported here was designed to provide an initial trial of the factor analytic technique on an important problem. The following data were collected for a sample of 240 urban places with populations of 10,000 persons or more. The source used was volume two of the U.S. census of population, 1950.⁴

<i>TV</i>	The proportion of households in the area which were equipped with television receivers, as of April, 1950.
<i>Y</i>	The median 1949 income, in thousands of dollars, received by families in the area.
$Y_1, \dots, Y_{14}$ ($X_1, \dots, X_{14}$)	The proportion of families in the area who had 1949 incomes falling into each of the 14 ranges reported by the census—the class limits are given in the tables and charts, below.

⁴ U.S. Bureau of the Census, *Seventeenth Census of the United States: 1950 Population*, Vol. II.

E The median number of years of formal education completed by males in the population.

$E_{11}, \dots, E_9$
($X_{15}, \dots, X_{23}$) The proportion of males in the population whose educational achievement placed them in each of the nine reported educational ranges—the class limits are given in the tables and charts, below.

Each variable is designated by a mnemonic symbol (for example, Y). In addition, the equivalent X_i notation will be used in the mathematical equations where convenient.

Each of the areas in the sample was identified with a source of television signals; all of them were selected so as to be within the primary coverage area of the stations operating out of an important metropolitan center.

Three television coverage variables were associated with each of the sample areas:

A The age, in months, of the oldest station received by families in the area, as of April, 1950.
(X_{24})

S The number of stations being received in April, 1950.
(X_{25})

W The length of time that live network programs (via coaxial cable or microwave) were available in the area, measured as a proportion of the age of the oldest station.
(X_{26})

Broadcasting-Telecasting Magazine and other TV market data books provided the source for these figures.⁵

The original sample contained urban places drawn from the primary television coverage areas of 41 metropolitan centers. Subsequently, a subsample of 141 urban places was selected for special treatment. They represented all of the observations in the sample for Boston, Los Angeles, Chicago, New York, Philadelphia, and Pittsburgh. Most of the factor analytic results reported in this paper are based upon these 141 areas.

The television ownership variable was subjected to a "logit" transformation to insure that the results reported here could be compared with those of Dernburg. This transformation is appropriate where the hypothetical relationship between dependent and independent variables is not linear, but rather is best described by a cumulative normal curve. The normal integral can be approximated by a logistic, or "Pearl-Reed," growth curve:

$$TV = \frac{M}{1 + e^{-(b_0 + b_1 X_1 + b_2 X_2 + \dots)}}$$

⁵ The methods for selecting sample areas and determining coverage were described in Massy, *op. cit.*, chap. iii, although data on counties rather than urban places were used there. (The technique of factor analysis was not used in the thesis.)

The parameter M is the hypothetical upper limit of growth, and its value is usually assumed at the start of the analysis.⁶ The logistic equation can be algebraically reduced to a form that permits linear regression on the independent variables:

$$\log_e \left(\frac{TV}{M - TV} \right) = b_0 + b_1X_1 + b_2X_2 + \dots$$

The variable on the left is called the "logit" of TV ownership and shall be denoted as L_{TV} . For the purposes of this study, the upper limit, M , is assumed to be one, or 100 percent ownership. Dernburg's argument about the desirability of the logit hypotheses for studying TV ownership is convincing.⁷

ESTIMATION OF OWNERSHIP CHARACTERISTICS FOR A GROUP OF SELECTED CITIES

The proportion of families falling into any cell in the income and education distributions will be correlated with those for other cells, as can be seen by a glance at Table 1. This table gives the simple correlation matrix for all variables, as obtained from the combined data for the 141 areas associated with the six metropolitan areas specified earlier.

Many of these correlations are very high, which prevents using a multiple regression of L_{TV} directly upon any substantial portion of the distribution data. If this were attempted the regression plane would be unstable—the standard errors of the coefficients would be very large—because of the high degree of collinearity between the independent variables. (Regression on all of the distribution variables would be impossible, since both the income-income and education-education submatrices in Table 1 are singular; that is, all income and education proportions must add up to one.)

Since direct multiple regression methods are ruled out, it was necessary to use an indirect method for separating the effects due to the various income and education classes. First, the raw data on all of the explanatory variables were factor analyzed in order to determine the orthogonal dimensions along which the urban places in the sample could be differentiated. (Factor analysis was discussed in "Statistical Analysis of Relations between Variables," elsewhere in this volume, and will not be described here.) Each area's factor scores were computed from the resulting matrix of factor loadings; these scores represent the value of each factor (that is, the position on each orthogonal dimension) for the area in question. Television ownership was related to the factors by means of a multiple regression of L_{TV} on the factor scores; multicollinearity was not a prob-

⁶ See, however, William F. Massy, *op. cit.*, pp. 127-30, for a maximum likelihood method of estimating M .

⁷ Dernburg, *op. cit.*, Appendix C.

TABLE 1
SIMPLE CORRELATION MATRIX
INCOME AND EDUCATION DISTRIBUTIONS, TV COVERAGE, AND OWNERSHIP*

Income Distribution Variables														Education Distribution Variables									Television Variables						
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)	(22)	(23)	(24)	(25)	(26)	(27)			
INCOME DISTRIBUTION (thousands of dollars)																													
0-0.5	(1)	100	37	31	45	53	40	27	-07	-28	-44	-49	-54	-37	-18	25	35	37	32	10	01	-24	-29	-19	-18	-13	23	-17	
0.5-1.0	(2)	100	82	76	59	36	25	22	-11	-16	-46	-52	-56	-36	20	24	23	25	07	11	-03	-05	-31	-23	-04	-15	-41		
1.0-1.5	(3)	100	75	67	41	33	25	11	-11	-45	-57	-62	-40	19	22	25	28	-01	19	-06	00	-34	-30	-10	-05	-43			
1.5-2.0	(4)	100	73	47	35	27	-08	-16	-52	-64	-70	-41	30	33	35	40	07	29	-12	-21	-41	-31	-16	10	-35				
2.0-2.5	(5)	100	73	47	35	27	-08	-16	-52	-64	-70	-41	30	33	35	40	07	29	-12	-21	-41	-31	-16	10	-35				
2.5-3.0	(6)	100	71	64	32	-04	-22	-57	-74	-77	-56	55	62	66	60	22	39	35	-56	-57	-50	-51	45	-43					
3.0-3.5	(7)	100	67	26	-05	-23	-42	-62	-69	-56	55	62	66	60	22	39	35	-56	-57	-50	-51	45	-43						
3.5-4.0	(8)	100	53	34	02	-25	-47	-73	-72	54	64	74	65	42	58	33	-64	-80	-37	-36	36	-23							
4.0-4.5	(9)	100	74	57	22	-10	-58	-79	20	21	27	36	43	67	08	-31	-77	-09	-07	-05	-01								
4.5-5.0	(10)	100	71	55	30	-23	-65	04	02	11	22	44	50	17	-21	-60	14	11	-17	18									
5.0-6.0	(11)	100	66	44	-38	-45	-19	-23	-14	-02	26	42	28	05	-34	17	15	-26	17										
6.0-7.0	(12)	100	72	36	-16	-38	-43	-39	-25	21	14	32	22	-04	40	30	31	40											
7.0-10.0	(13)	100	65	16	-38	-48	-49	-46	-04	-16	31	39	27	42	37	-36	37												
more than 10.0	(14)	100	64	-35	-40	-51	-54	-49	-68	04	45	90	24	21	-11	19													
EDUCATION DISTRIBUTION (formal education completed)																													
None	(15)	100	74	67	52	20	23	54	55	-51	-33	-33	32	-27															
Elementary: 1-4	(16)	100	89	71	29	20	-69	-67	-59	-32	-40	30	-37																
Elementary: 5-6	(17)	100	80	39	32	-68	-75	-67	-32	-40	30	-37																	
Elementary: 7	(18)	100	51	39	-64	-69	-70	-16	-23	29	-08	10	-01	00	10	-01	00	10	-01	00	10	-01	00	10	-01	00			
Elementary: 8	(19)	100	18	44	-57	-60	10	-01	00	10	-01	00	10	-01	00	10	-01	00	10	-01	00	10	-01	00	10	-01	00		
High school 1-3	(20)	100	08	-41	-71	-17	-15	19	09	-04	15	10	09	-04	15	10	09	-04	15	10	09	-04	15	10	09	-04	15	10	
High school 4	(21)	100	42	25	-04	15	10	09	-04	15	10	09	-04	15	10	09	-04	15	10	09	-04	15	10	09	-04	15	10		
College 1-3	(22)	100	60	32	38	-53	10	32	38	-53	10	32	38	-53	10	32	38	-53	10	32	38	-53	10	32	38	-53	10		
College 4	(23)	100	21	23	-13	13	21	23	-13	13	21	23	-13	13	21	23	-13	13	21	23	-13	13	21	23	-13	13			
Months of TV coverage	(24)	100	85	-70	76	100	85	-70	76	100	85	-70	76	100	85	-70	76	100	85	-70	76	100	85	-70	76	100	85	-70	76
Number of stations	(25)	100	71	63	100	71	63	100	71	63	100	71	63	100	71	63	100	71	63	100	71	63	100	71	63	100	71	63	
Percent network coverage	(26)	100	71	63	100	71	63	100	71	63	100	71	63	100	71	63	100	71	63	100	71	63	100	71	63	100	71	63	
"Logit" of TV ownership	(27)	100	71	63	100	71	63	100	71	63	100	71	63	100	71	63	100	71	63	100	71	63	100	71	63	100	71	63	

* Decimal points omitted.

lem, since the computed scores were nearly uncorrelated. Finally, the relative percentage ownership of each subgroup in the income and education distribution was computed by taking a weighted average of the regression coefficients for each factor, with the square of the factor loading of each variable on that factor as weight. The steps will be explained in detail later in this section.

The overall procedure is summed up in the flow diagram presented in Figure 1. The lightly bordered boxes refer to raw data, and intermediate and final results; all of this information was stored on punched cards. The heavily bordered boxes denote computer programs for processing the information. All programs except the regression package were written by the author and run on the IBM 7090 computer at the M.I.T. Computation Center. The factor loadings were computed and rotated according to the methods of Holzinger⁸ and Kaiser⁹; respectively. The program can extract and rotate 10 factors from observations on 26 variables in close to one minute. Regressions and some preliminary data manipulation (including the logit transformation on TV ownership) were performed on the IBM 1620 machine at the M.I.T. School of Industrial Management's computer facility.

Table 2 gives the rotated loadings obtained from factoring the 26 income, education, and coverage variables for the 141 observations whose correlation coefficients were presented in Table 1. Ten factors were rotated. The five most important ones may be interpreted as follows:

- FACTOR 1: Clearly an *education* dimension, it is strongly positive on low, neutral on middle, and strongly negative on high education classes. None of the income or coverage variables have high enough loadings to be important.
- FACTOR 2: This dimension brings out the communality between the *income* and *education* characteristics of middle and high social groups in the sample.
- FACTOR 3: An *income* factor, this dimension is positively loaded on the income classes below \$2,500 and negative on those above \$5,000.
- FACTOR 4: This dimension picks up the television *coverage variables*. They are strongly related to each other, but are not associated with any particular income or education class.
- FACTOR 5: This factor separates out the *lowest income* group as a distinct dimension.

The other factors are fairly specific to particular variables and can be easily interpreted by the reader.

Factor scores were obtained from the factor loadings matrix and the raw values of the explanatory variables. If the number of factors extracted had been equal to the number of variables factored (that is,

⁸ Karl J. Holzinger, Harry H. Harman, *Factor Analysis* (Chicago, Ill.: University of Chicago Press, 1941).

⁹ Henry F. Kaiser, "The Varimax Criterion for Analytic Rotation in Factor Analysis," *Psychometrika*, Vol XXIII (1958), pp. 187-200.

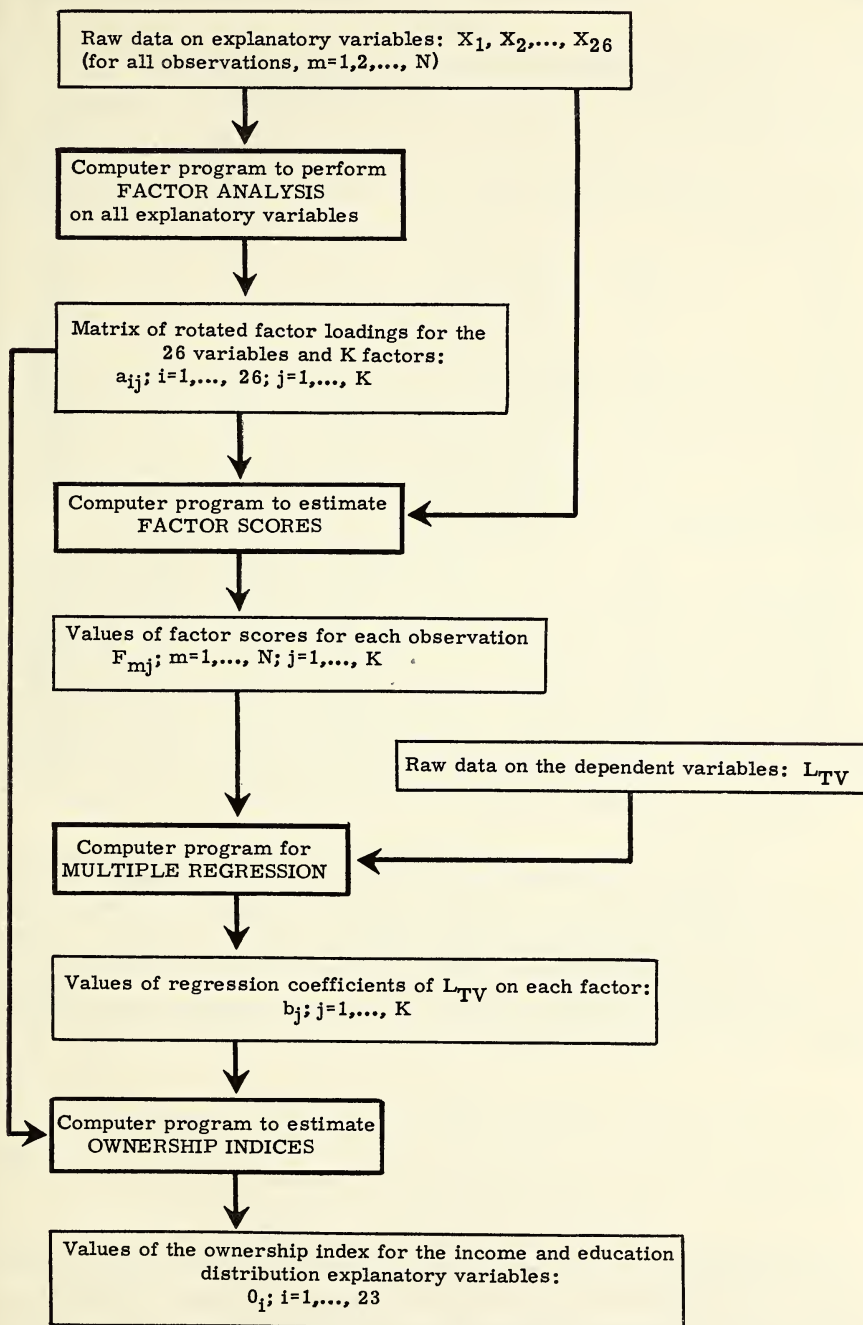


FIGURE 1. Flow diagram for estimation procedures.

TABLE 2
FACTOR LOADINGS FOR TV COVERAGE VARIABLES AND INCOME AND
EDUCATION DISTRIBUTIONS, FIVE SELECTED CITIES

	Factors†										Proportion of Variance Extracted
	1	2	3	4	5	6	7	8	9	10	
Income Distribution (thousands of dollars)											
0-0.5.....	21	-15	34	-07	87*	57*	-08	03	07	01	0.948
0.5-1.0.....	04	05	91*	05	09	49*	-01	-16	04	07	0.882
1.0-1.5.....	06	09	92*	-07	-01	-08	00	-07	12	-06	0.887
1.5-2.0.....	13	10	88*	-09	17	01	-02	21	-06	09	0.889
2.0-2.5.....	42	18	66*	-36	27	02	-14	08	10	11	0.884
2.5-3.0.....	43	13	37	-40	14	07	-05	15	59*	11	0.912
3.0-3.5.....	48	52*	20	-26	12	07	-31*	02	40*	09	0.897
3.5-4.0.....	06	89*	22	-01	-09	11	-13	02	01	03	0.877
4.0-4.5.....	00	90*	-18	12	-08	14	01	-09	-11	02	0.903
4.5-5.0.....	-15	73*	-14	11	-28	04	35*	07	-23	-14	0.869
5.0-6.0.....	-29	43	-51*	24	-17	16	49*	15	02	-11	0.901
6.0-7.0.....	-31	13	-62*	28	-17	-03	48*	-17	-13	05	0.875
7.0-10.0.....	-35	-42	-63*	15	-05	-13	23	-21	-20	-09	0.881
More than 10.0.....	-23	-82*	-30	15	-17	-11	-18	16	-25	-05	0.975
Education Distribution (formal education completed)											
None.....	60*	10	17	-20	02	-01	-04	04	09	71*	0.955
Elementary: 1-4.....	85*	12	17	-21	14	-04	-13	-10	07	25*	0.923
Elementary: 5-6.....	85*	22	18	-24	17	04	-12	01	11	10	0.928
Elementary: 7.....	80*	28	24	-12	11	19	-02	21	04	-12	0.894
Elementary: 8.....	35	37	-01	08	05	84*	04	-05	04	-03	0.980
High school: 1-3.....	10	72*	13	-11	01	-14	-04	54*	20	04	0.916
High school: 4.....	-89*	27	-06	-07	06	-23	-40	11	-04	-03	0.919
College: 1-3.....	-55*	-28	03	35	-22	-39	24	-28	00	-21	0.888
College: 4.....	-45*	-76*	-27	10	-11	-17	-01	-07	-19	-12	0.951
Months of TV coverage....	-02	-04	-27	92*	-01	07	05	08	-06	-12	0.944
Number of stations.....	-23	-02	-07	92*	04	02	-01	07	-07	04	0.915
Percent network coverage..	15	-07	-11	-83*	18	06	-14	33*	04	07	0.899

* Loadings that are high enough to be of interest.

† Decimal points omitted.

$K = 26$), factor scores could have been calculated by straightforward solution of the 26 simultaneous linear equations relating the scores to the original variables, in which the factor loadings appear as constants. That is, we could solve for the F 's in system (1), where the X 's are the original (independent) variables, and the a 's are the computed factor loadings.¹⁰

$$\begin{aligned}
 X_1 &= a_{11}F_1 + a_{12}F_2 + \cdots + a_{1K}F_K \\
 &\vdots \\
 X_{26} &= a_{26,1}F_1 + a_{26,2}F_2 + \cdots + a_{26,K}F_K
 \end{aligned} \tag{1}$$

Unfortunately, the number of factors extracted was not equal to the number of variables; one of the major purposes of factor analysis is to summarize the information contained in a particular set of variables by means of a *smaller* number of factors. Equation system (1) will be over-

¹⁰ System (1) is the fundamental equation set for factor analysis; it is discussed in "Statistical Analysis of Relations Between Variables," elsewhere in this book.

determined when K is less than the number of variables, since there will be more equations than factor scores to be determined. The factor scores can then be approximated by means of least squares, that is, we can regard the F 's as the coefficients of an equation for predicting the value of any X_i , given the values of the a_{ij} ($j = 1, \dots, K$). This amounts to estimating the coefficients of the following "multiple regression equation" for each observation (m) in the sample:

$$X_{im} = F_{1m}a_{i1} + F_{2m}a_{i2} + \dots + F_{Km}a_{iK} + u_{im}; \\ i = 1, \dots, 26 \quad (2)$$

The X 's and a 's are considered to be known numbers, with the F 's being regression coefficients. The estimates of F_1 through F_K are computed by minimizing the sum of squares of the u_i over the 26 "data points" available for each sample area. The regression equation (2) is solved for each sample area separately, and the resulting estimates of the F_j are the factor scores for that area. The mathematical derivation of this procedure is given in the appendix to this paper.

The estimated factor scores were used as explanatory variables in a regression with logit TV as the dependent variable. Strictly speaking, this regression contains errors in the independent variables as well as in the equation, but this problem had to be neglected because estimates of the variances of the factor scores were not available.¹¹ Standard least squares procedures yielded the following regression equation, based on factor scores calculated from the loadings of Table 2:

$$L_{TV} = -2.42F_1^* + 4.04F_2^* - 2.88F_3^* - 1.23F_4^* - 2.26F_5^* + 0.15F_6 \\ (0.66) \quad (0.95) \quad (1.41) \quad (0.34) \quad (1.09) \quad (0.43) \\ -0.33F_7 + 3.24F_8^* - 0.32F_9 - 0.53F_{10} \quad R^2 = 0.84 \\ (0.79) \quad (0.46) \quad (0.69) \quad (0.49)$$

The standard errors of the coefficients are given within the parentheses; the starred values are significant by the t test at the 0.05 level of significance.

The coefficients of factors 1 through 5, and 8, are significant at the 5 percent level or beyond. Another regression, containing only the significant explanatory variables, was computed with the following results:

$$L_{TV} = -2.09F_1^* + 3.83F_2^* - 3.54F_3^* - 1.30F_4^* - 2.78F_5^* + 3.29F_8^* \quad R^2 = 0.83 \\ (0.41) \quad (0.80) \quad (1.20) \quad (0.18) \quad (0.97) \quad (0.43) \quad (3)$$

All variables are significant, and the proportion of variance explained by the regression is only slightly smaller than in the preceding equation. The

¹¹ It may prove to be possible to obtain approximations to the variances of the factor scores as a byproduct of the least squares reduction procedure described above, perhaps under the (questionable) assumption that the a_{ij} are estimated without error.

values of the coefficients were changed somewhat, but not by a serious amount. (The least squares approximation for obtaining the factor scores introduces correlation between what should, strictly speaking, be orthogonal factors.)

Equation (3) can best be interpreted if we compute the standardized regression coefficients:

$$\beta_i = \frac{\sigma_i}{\sigma_{L_{TV}}} b_i$$

The β 's do not depend upon the scale of measurement for any of the variables, and are useful for assessing the effects of equally likely displacements of the various factors. For (3), they are:

$$L_{TV} = -5.25F_1 + 2.64F_2 - 4.14F_3 - 14.0F_4 - 5.18F_5 + 23.9F_8 \quad (4)$$

The coefficients of F_8 and F_4 are the largest in absolute value. While this was not suspected at the outset, both evidently relate to the network coverage variable: the sign of the coverage loading on factor 4 is negative, as is β_4 , while both $a_{26,8}$ and β_8 are positive. Four of the six metropolitan centers in the sample, including 107 of the 141 observations, had a market age of exactly 60 months, which substantially reduced the effect of this variable. This fact, coupled with the β coefficients themselves, suggests that length of network coverage was substantially more important in determining the level of television ownership than was the number of stations (which ranged from one to seven in the sample). This inference is borne out by the regressions to be reported in Table 3.

Turning to the coefficients for the other factors, we see by β_1 that medium to high education, and high income to some extent, are strongly associated with TV ownership while low education stands in an inverse relationship. β_5 shows conclusively that extremely low income groups were not likely to be TV owners, while β_3 suggests that the positive relationship between income and ownership extends most of the way up the income distribution. This conclusion must be somewhat modified because of β_2 : the middle income and education classes were strongly TV oriented, but the relationship is inverted at the highest levels in both distributions.

While the regression coefficients can be interpreted directly in terms of the factor loadings on the original variables, as was done above, the conclusions are highly subjective at best. A more straightforward way of relating them to the original variables is desirable. The following method seems to be an appropriate way of accomplishing this purpose, although no airtight theoretical justification of its validity is available.

The value of an *ownership index*, showing the relative contribution of each of the income and education subgroups to aggregate television

ownership, was calculated as the weighted average of the β coefficients for all of the factors. The squares of the factor loadings were used as weights. The formula for the ownership index of variable 1 (that is, families in the lowest income class) is:

$$O_1 = a_{i1}^2 \beta_1 + a_{i2}^2 \beta_2 + \dots + a_{iK}^2 \beta_K$$

Index values for all of the other income and education subgroups were calculated in a similar manner, and are plotted as a profile in Figure 2

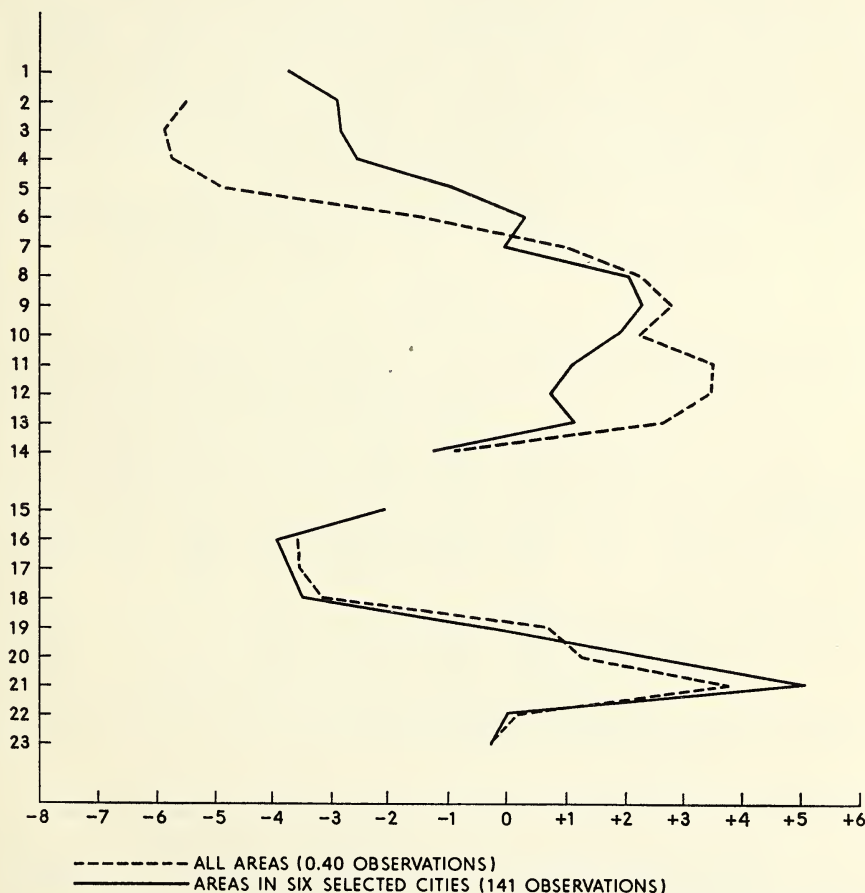


FIGURE 2. Ownership index profiles: income and education distributions for all sample areas and five selected cities (10 factors).

(solid line). Squares of the factor loadings were used as weights because these quantities represent the proportion of the variance of each of the X variable that is accounted for by the factor in question.

The *ownership index profile* for areas in the six selected cities indicates that, for the sample, television ownership increased rather steadily

up to an income level of from four to five thousand dollars, and then declined sharply in the extremely high income range (over \$10,000). With the exception of a relatively small upswing in the lowest class, education was associated with increasing ownership up to the level of high school graduation. College-trained persons tended to be much less likely to own television than their less educated neighbors.

COMPARISON WITH DERNBURG'S RESULTS

An ordinary regression of L_{TV} upon median income and education, and the television coverage variables, was performed for the full sample of 240 urban places and compared with Dernburg's results. This was done for two reasons: (1) to see if his findings could be replicated; and (2) to determine whether the present sample was sufficiently like his to allow a later comparison of the factor analytic-regression results with his straight regression-analysis of variance conclusions.

The answer to both questions was "yes," as can be seen from the regression coefficients presented in Table 3. The first two rows are co-

TABLE 3

REGRESSION COEFFICIENTS

"Logit" Television Ownership on Median Income and Education and TV Coverage Variables

	Con- stant	Median Income	Square of Median Income	Number of Months of TV Coverage	Number of Stations	Proportion Network Coverage	Median Education Level	R ² (n)
Massy (1962):								
With education..... <i>b</i> (<i>b/as</i>)	-4.58	+0.461 (3.1)	-0.034 (2.6)	+0.034 (12.3)	+0.048 (2.1)	+1.431 (12.3)	-0.0006 (0.02)	0.817 (240)
Without education..... <i>b</i> (<i>b/as</i>)	-4.58	+0.460 (3.5)	-0.034 (2.8)	+0.034 (13.3)	+0.048 (2.3)	+1.430 (12.3)		0.817 (240)
Dernburg (1958):*								
Equation I: <i>b</i> <i>b</i> (<i>b/as</i>)	-4.50	+0.622 (20.7)	-0.044 (12.6)	+0.031 (11.5)	+0.037 (4.5)			0.695 (1373)
Equation III: <i>b</i> <i>b</i> (<i>b/as</i>)	-3.61	+0.72 (†)	-0.058 (†)	+0.056 (†)	-0.004 (†)			0.486 (1400)

* Source: Thomas F. Dernburg, *Studies in Household Economic Behavior*, Yale Studies in Economics, Vol. IX (New Haven, Conn.: Yale University Press, 1958), Tables 1 and 2.

† Not given by Dernburg.

efficients of the regression of logit TV upon median income, median income squared, the television coverage variables, and (in the first case only) median education. The second two are those equations from Dernburg which are most directly comparable to these findings; equation I:*b* refers to cities with only local TV coverage, while III:*b* refers to those with both local and distant stations available. Dernburg handled the education variables in his analysis of regression residuals, and therefore an education coefficient is not available; he did not use a network coverage variable. Differences due to the use of family income in the present sample and income of families and unrelated individuals in Dernburg's may have affected the findings somewhat.

Almost all of the coefficients for Dernburg's equation 1:*b* agree fairly well with the ones presented here, although their *t* ratios are larger because of the difference in sample size. (The sample used for this paper corresponds most closely to Dernburg's Class I areas.) Some difference would be expected because of the different aggregation characteristics of relatively small census tracts and relatively large urban places.

An ownership index profile was also computed for the 240 observations in the full sample; it is presented as the dashed line in Table 2. The profile indicates that increasing income leads to increasing ownership up to levels of about \$7,000. This checks closely with Dernburg's result of "\$6,500 to \$7,500." Ownership reaches a sharp peak at 12 years of formal education—high school graduation—and declines sharply thereafter. Dernburg's peak was in the 10.0–11.9 years class, with an equally sharp drop off at higher levels.¹² His decrease of ownership on the low side of high school graduation was much less pronounced than mine—a point that needs future investigation.

OWNERSHIP PROFILES FOR FIVE CITIES SEPARATELY

Ownership index profiles were calculated for each of five metropolitan areas separately, in an effort to learn whether the results in Table 2 would be stable for different geographic areas and much smaller sample sizes. Two different procedures were followed.

Figure 3 presents a family of profiles based upon a common factor matrix, obtained for the five metropolitan centers as a group. The regression and β coefficients from which they were computed are given in Table 5, and the factor loadings matrix is presented in Table 4. Only six factors were rotated; they are identified as follows:

FACTOR 1: This is primarily an education dimension, although there is some significant communality with the middle income classes.

FACTOR 2: A common dimension over income and education, it picks up the middle and high classes in both distributions.

FACTOR 3: This factor represents the income dimension; it is heavily loaded on the medium low, and medium high income levels.

FACTOR 4: Another education factor, this one is strongly positive on seventh and eighth grade, as opposed to high school and early college, educational achievement.

FACTOR 5: This dimension is specific to the very low income class.

FACTOR 6: Another common dimension on education and income; it is most heavily loaded on the medium high income classes and the "some college" group, while being negative on the "some high school" and middle income groups.

Factors 1 through 3, and 5, correspond to their counterparts in Table 2. The differences between factors are not as sharply delineated as was the case there, undoubtedly because fewer factors were rotated. Factors 4

¹² Dernburg, *op. cit.*, Figure 8.

and 6 represent a combination of factors 6-10 of Table 2, with the effects of the TV coverage variables removed.

The regression coefficients in Table 5 show a considerable amount of variation between metropolitan areas, but since most of them are not significant at the 0.05 level, conclusions must be drawn with caution.

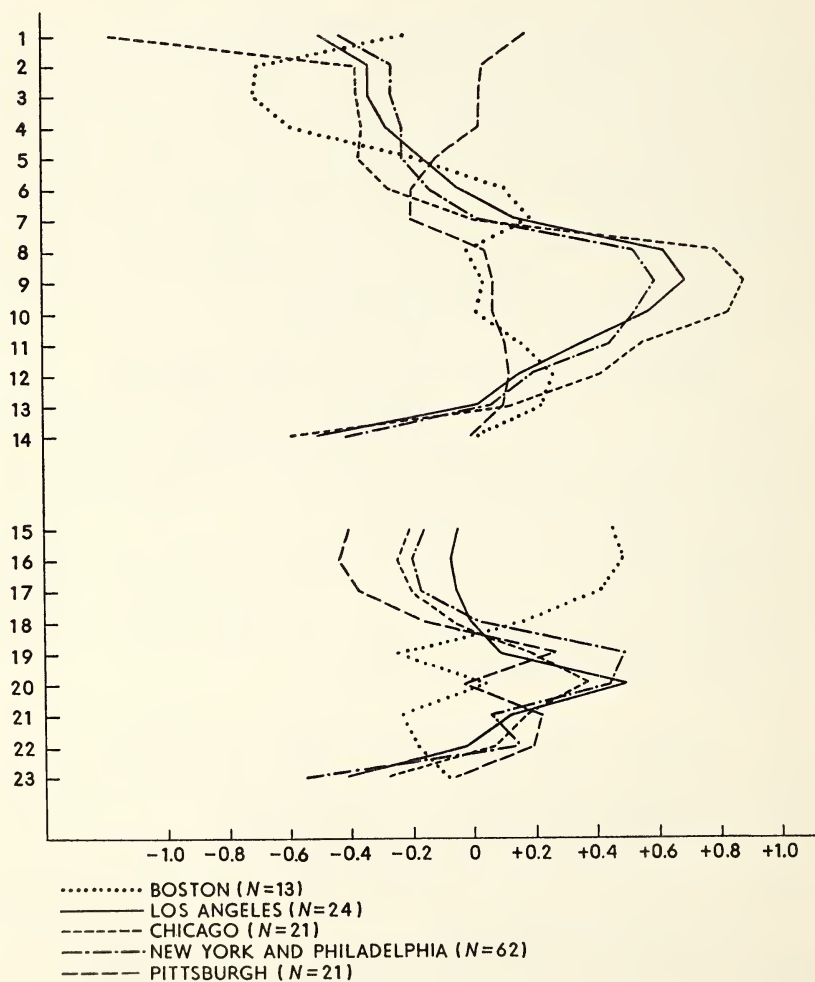


FIGURE 3. Ownership index profile income and education distributions: common factors for five selected cities.

The β coefficients for factor 2, income and education, are significant for Los Angeles and New York, and almost so for Boston: we note that their values range from +0.74 to +1.08, indicating a fairly stable positive association of ownership with income of from \$3,500 to \$6,000, and an incomplected high school education. The same arguments hold for factor 3; the relationship of television ownership is negative on the \$500

TABLE 4
FACTOR LOADINGS FOR INCOME AND EDUCATION DISTRIBUTIONS,
FIVE SELECTED CITIES

	Factor†						Proportion of Variance Extracted
	1	2	3	4	5	6	
Income distribution (thousands of dollars)							
0-0.5	-21	-20	34	10	83*	-11	0.915
0.5-1.0	-07	04	92*	04	12	10	0.869
1.0-1.5	-08	08	93*	-05	-00	-02	0.886
1.5-2.0	-16	08	86*	02	15	-17	0.816
2.0-2.5	-51	12	65*	01	27	-31*	0.861
2.5-3.0	-58	13	36	-04	23	-44*	0.723
3.0-3.5	-60	46	23	06	15	-40*	0.800
3.5-4.0	-12	85*	23	10	-07	-15	0.824
4.0-4.5	-02	88*	-16	15	-08	-08	0.836
4.5-5.0	22	78*	-15	10	-29	17	0.802
5.0-6.0	35	55*	-52*	13	-12	33*	0.829
6.0-7.0	32	23	-61*	-06	-15	55*	0.861
7.0-10.0	41	-37	-62*	-13	-07	40*	0.867
More than 10.0	33	-83*	-31	-08	-23	06	0.945
Education distribution (for- mal education completed)							
None	-87*	07	15	-14	06	02	0.808
Elementary: 1-4	-92*	02	17	09	10	-09	0.895
Elementary: 5-6	-86*	13	18	20	13	-21	0.893
Elementary: 7	-67*	22	23	43*	04	-25	0.803
Elementary: 8	-25	39	02	83*	12	-01	0.910
High school: 1-3	-17	73*	11	-15	-01	-47	0.810
High school: 4	75*	27	-06	-48*	12	-07	0.887
College: 1-3	62*	-22	06	-34*	-22	47*	0.821
College: 4	51*	-73*	-26	-19	-13	16	0.948

* Loadings that are high enough to be of interest.

† Decimal points omitted.

to \$2,500 income groups, and positive on the \$5,000 to \$10,000 classes. This relationship is not as strong as the one for factor 2. The values of β_5 show a significant negative relationship between ownership and membership in the extreme low income group for Los Angeles, Chicago, and New York, although the effect is almost three times as strong for Chicago as for the other cities.

Boston presents the only anomaly; the coefficient for education is almost significant and has the same sign as the low education factors in Table 4. On the surface, this would indicate that the lower educational groups purchased television to a greater extent than did the higher ones. Factor 1 is also negatively loaded on the \$2,000-\$3,500 income groups: it is possible that their effect swamped that of education, but this seems unlikely because (1) the education variables are more heavily loaded on the factor than is income; and (2) the variances within the education distribution are on the average larger than those of the middle and upper segments of the income distribution. We therefore cannot rule out the

TABLE 5
REGRESSION COEFFICIENTS OF LOGIT TV ON INCOME AND EDUCATION
FACTOR SCORES FOR FIVE SELECTED CITIES

City	Factor						R^2 (d.f.)
	1	2	3	4	5	6	
Boston							
b	-5.14	+0.14	-15.37	-2.27	-1.62	-0.08	0.887
σ_b	(2.18)	(3.80)	(6.45)	(1.38)	(4.47)	(2.17)	(6)
β	-0.62	+0.02	-0.83	-0.43	-0.18	-0.10	
Los Angeles							
b	+2.31	+8.24*	-9.66*	-0.50	-8.78*	-3.29	0.754
σ_b	(4.20)	(2.04)	(3.36)	(1.76)	(3.61)	(2.42)	(17)
β	+0.09	+0.87	-0.40	-0.05	-0.53	-0.22	
Chicago							
b	+2.17	+5.93	-16.58	+4.30	-12.28*	+5.45	0.530
σ_b	(2.61)	(3.72)	(8.06)	(2.57)	(4.49)	(2.85)	(14)
β							
New York and Philadelphia							
b	+2.30	+4.06*	-6.82*	+0.96	-3.25*	-0.75	0.304
σ_b	(1.34)	(1.35)	(3.03)	(0.74)	(1.32)	(1.31)	(55)
β	+0.28	+0.74	-0.31	+0.16	-0.53	-0.01	
Pittsburgh							
b	+4.42	+2.27	-0.61	+3.46	+4.46	+2.81	0.447
σ_b	(9.44)	(2.17)	(8.04)	(3.85)	(3.47)	(2.69)	(14)
β	+0.54	+0.08	+0.03	+0.42	+0.29	+0.23	

* Significant at the 0.05 level on a two-tailed test.

possibility that Boston is structurally different from the rest of the sample as far as the effect of education on television ownership is concerned.

Figure 3 presents the ownership index profiles calculated from the regression coefficients of Table 5. As would be expected, the profiles for Los Angeles, Chicago, and New York fall very close together. Those for Boston and Pittsburgh are distributed erratically, undoubtedly as a result of the instability of the regression parameters. The inversion of the education effect for Boston, which was noted in connection with the β coefficients, is carried through to the ownership index profiles.

An experiment with individual factor analyses of data for the five cities taken separately was also attempted. Figure 4 reports ownership index profiles based on regressions of L_{TV} upon scores computed from the area's own set of factors, rather than from a single factor analysis for all the metropolitan areas, as was the case above. The profiles are considerably less stable than those of Figure 3. They are again based on regressions upon six factor scores, which had the following significance characteristics. (The factors themselves represented somewhat different dimensions for each city, and it would not be worthwhile to report them all here.)

The profiles for Los Angeles and New York are very much like the

ones in Figure 3; this is reasonable considering that these two cities account for 86 out of the 141 observations in the sample. Chicago's income and education profiles are both skewed toward the higher classes. We may take some stock in this result because (1) the shift is quite marked, and (2) the regression coefficients for Chicago were more significant

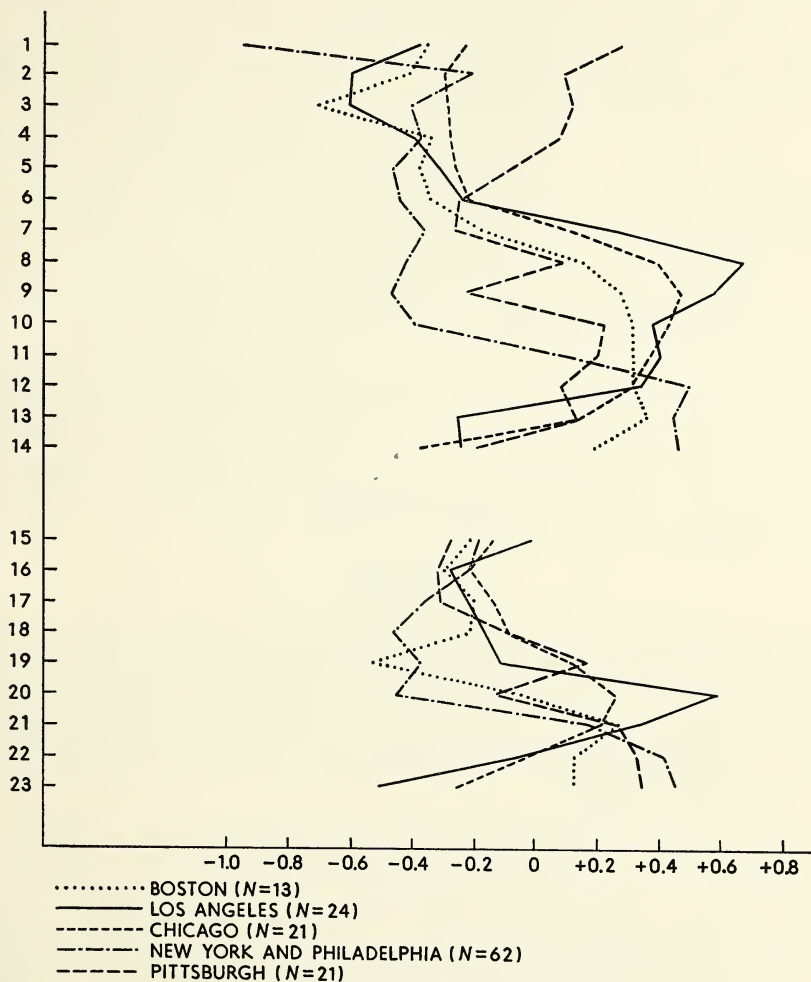


FIGURE 4. Ownership index profile income and education distributions: separate factors for five selected cities.

than those of all the areas except Los Angeles. Pittsburgh's profile cannot be considered seriously, because the underlying regression coefficients are unstable. Boston must be questioned for this reason as well, although it is interesting to note that the positive values for the low education indices, noted earlier, are not reproduced in Figure 4.

In retrospect, it seems that six factors were too few to provide for an

adequate comparison between cities, based upon separate factor analyses. The exact shape of the profiles is dependent upon the particular factors extracted, as well as upon the regression coefficients. In a word, the discriminating power of the analysis declines as the number of factors declines. The results are "lumpy" for small numbers of factors; even though the regression procedure insures the best possible fit between L_{TV} and the factor scores, this is not sufficient to overcome some kinds of differences in the factor loadings matrix. One would not expect to dif-

City	Total Number of Coefficients	Cumulative Number Significant at the Level of		
		0.01	0.05	0.10
Boston.....	6	1	1	4
Los Angeles.....	6	3	5	5
Chicago.....	6	1	3	4
New York and Philadelphia.....	6	1	3	3
Pittsburgh.....	6	0	0	2

ferentiate between two income and education classes that are both heavily loaded on the same factors, for example. Increasing the number of factors increases the specificity of each, though at the cost of degrees of freedom in the regression, and thus increases the sensitivity of the profiles (although not their statistical stability).

SUMMARY

Factor analysis was used to define independent dimensions within the 23 subgroups of the income and education distributions for a sample of 141 urban places in and around six major metropolitan areas. Factor scores were computed by a least squares method, and used as explanatory variables for regressions aimed at predicting the "logit" of the percentages of television ownership. The β coefficients from these regressions were in turn transformed into ownership index profiles.

The relationship between income and television ownership was found to be nonlinear. Peak ownership occurred at incomes of from \$4,000 to \$4,500 for the sample of 141 areas, and at from \$6,000 to \$7,000 for a larger sample of 240 areas drawn from 47 metropolitan centers. The latter result agrees closely with Dernburg's findings. The profiles for the education distribution showed a similar pattern. The peak of ownership occurred at the level of high school graduation for both samples, slightly higher than in Dernburg's study. Analysis of areas in five metropolitan centers, taken individually, produced roughly similar results.

The methods utilized in this preliminary study appear to be very promising in situations where a large number of interrelated explanatory variables must be employed. While additional work is needed to refine the techniques (particularly with respect to questions of statistical significance), there seem to be no important barriers to their implementation.

APPENDIX: THE DERIVATION OF FACTOR SCORES

The derivation of factor scores will be demonstrated for the case of three variables and two factors. It is easily generalized to problems involving n

variables and k factors. (See the technical note at the end of this appendix.)

Factor analysis begins with the idea that each of the original variables (the X_i) can be set equal to a linear combination of the unknown factor scores (F_j) and a residual (the u_i). For the 3×2 case this is written:

$$\begin{cases} X_{1m} = a_{11} F_{1m} + a_{12} F_{2m} + u_{1m} \\ X_{2m} = a_{21} F_{1m} + a_{22} F_{2m} + u_{2m} \\ X_{3m} = a_{31} F_{1m} + a_{32} F_{2m} + u_{3m} \end{cases} \quad (A1)$$

where the subscript m refers to the observation number. The a 's are known as factor loadings; they are the output of the factor analysis procedure. By itself, factor analysis cannot estimate values for the factor scores, F_{1m} and F_{2m} : that is the question confronted in this appendix.

The system (A1) must be solved for F_{1m} and F_{2m} , in terms of the known a 's, and X_{1m} , X_{2m} , and X_{3m} . This must be done separately for each observation in the sample, since values of the F 's for each observation are desired. Recalling that all of the following calculations refer to the same observation, we will drop the subscript m in order to keep the notation compact.

The system (A1) is overdetermined and cannot be solved directly; it contains three equations but only two unknowns (F_1 and F_2). Two courses of action are open. First, we could take any subset of two equations and solve them. Taking the equations for X_1 and X_2 and neglecting the unknown u 's, we have:¹³

$$\begin{cases} F_1 = \frac{a_{22}X_1 - a_{12}X_2}{a_{11}a_{22} - a_{12}a_{21}} \\ F_2 = \frac{a_{11}X_2 - a_{21}X_1}{a_{11}a_{22} - a_{12}a_{21}} \end{cases} \quad (A2)$$

The second alternative takes all of the variables into account symmetrically, whereas the system (A2) does not bring the variable X_3 into solution. We proceed by minimizing the sum of the squares of the residuals.¹⁴ The approach is reasonable because factor scores which result in the best possible approximation to the X 's are desired. The sum of squares of the u 's is written:

$$S = \sum_{i=1}^3 u_i^2 = \sum_{i=1}^3 (X_i - a_{i1}F_1 - a_{i2}F_2)^2$$

It is minimized by differentiating partially with respect to each of the F 's, and setting the derivatives equal to zero:

¹³ This direct method of solution would be appropriate if the number of factors were equal to the number of variables. Then the number of unknown factors would be the same as the number of equations.

¹⁴ The residuals in (A1) are not homoscedastic. In future work, they can be normalized by dividing each equation through by the quantity $(1 - a_{i1}^2 - a_{i2}^2)$, which is proportional to σu_i^2 . The method used in this paper produces unbiased estimates of the factor scores, but fails to be fully efficient. After normalizing, this method is equivalent to that of "minimizing unique factors" due to M. S. Bartlett. It is discussed in Harry Harmon, *Modern Factor Analysis* (Chicago: University of Chicago Press, 1960), pp. 356-60.

$$\left. \begin{aligned} \frac{\partial S}{\partial F_1} &= -2\sum a_{i1}(X_i - a_{i1}F_1 - a_{i2}F_2) = 0 \\ \frac{\partial S}{\partial F_2} &= -2\sum a_{i2}(X_i - a_{i1}F_1 - a_{i2}F_2) = 0 \end{aligned} \right\} \quad (A3)$$

All summations are taken over i , from 1 to 3.

System (A3) can be reduced to:

$$\left. \begin{aligned} F_1\sum a_{i1}^2 + F_2\sum a_{i2}a_{i1} &= \sum a_{i1}X_i \\ F_2\sum a_{i1}a_{i2} + F_2\sum a_{i2}^2 &= \sum a_{i2}X_i \end{aligned} \right\} \quad (A4)$$

The reader may recognize (A4) as being identical in form to the normal equations of multiple regression which, in fact, are derived by minimizing $\sum u^2$ in a similar manner.¹⁵ It consists of two linear equations, contains two unknowns, and can be solved by ordinary methods. The equations for F_1 and F_2 are:

$$\left. \begin{aligned} F_1 &= \frac{(\sum a_{i2}^2)(\sum a_{i1}X_i) - (\sum a_{i1}a_{i2})(\sum a_{i2}X_i)}{(\sum a_{i1}^2)(\sum a_{i2}^2) - (\sum a_{i1}a_{i2})^2} \\ F_2 &= \frac{(\sum a_{i1}^2)(\sum a_{i2}X_i) - (\sum a_{i1}a_{i2})(\sum a_{i1}X_i)}{(\sum a_{i1}^2)(\sum a_{i2}^2) - (\sum a_{i1}a_{i2})^2} \end{aligned} \right\} \quad (A5)$$

We remind the reader that equations (A5) must be evaluated for each observation in the sample.

Technical Note

These ideas are much more compactly presented in the notation of matrix algebra.¹⁶ We observe the following definitions:

$\mathbf{A}$ = The factor loadings matrix.

$\mathbf{x}_m$ = The vector of original variables for the m^{th} observation in the sample.

$\mathbf{f}_m$ = The vector of factor scores for the m^{th} observation in the sample.

The elements of the vector of factor scores are computed from the following vector equation:

$$\mathbf{f}_m = (\mathbf{A}'\mathbf{A})^{-1} (\mathbf{A}'\mathbf{x}_m)$$

The first term is obtained by premultiplying the factor loadings matrix (which has dimension $n \times k$) by its transpose ($k \times n$) and inverting the $k \times k$ product matrix. This result is called the inverted "factor-factor covariation matrix." These operations can be performed once, at the beginning of the calculations for a given sample. Then the vector of original variables ($n \times 1$) is premultiplied by the factor loadings matrix transpose ($n \times k$) to produce a k element vector of factor-variable covariations. This operation is performed for each observation in the sample, and the result premultiplied by the inverse of the factor-factor covariation matrix to obtain the vector of factor loadings.

¹⁵ R. L. Anderson and T. A. Bancroft, *Statistical Theory in Research* (New York: McGraw-Hill Book Co., Inc., 1952), p. 169.

¹⁶ The notation and operations of matrix algebra are explained in highly readable fashion in Anderson and Bancroft, *ibid.*, pp. 172-77.

READINGS ON SIMULATION

PART FIVE



Simulation

COMPREHENSIVE REPORTS OF MARKETING simulations are only beginning to appear in the literature of the profession. Of the four papers presented in this section, two were published in 1961. Another is scheduled to be published in early 1963, and the fourth is an article-length report which appears in this volume for the first time.

The first paper, "The Simulmatics Project," by Pool and Abelson, reports a simulation project aimed at predicting likely voting behavior during the 1960 presidential election campaign. Techniques such as the simulation illustrated in this article may have a number of applications in marketing. The reader may find it useful to attempt to answer the following questions in the process of analyzing this article:

1. What are the characteristics of the voting problem that facilitate the use of simulation as a technique for analysis and prediction?
2. To what extent are these characteristics present in the analysis of consumer behavior with respect to various products (say the choice of a refrigerator, an automobile, a television program, and/or a stereo set)?

The Orcutt, Greenberger, Korbelt, and Rivlin paper entitled "A Stochastic Microanalytic Model of a Socioeconomic System," presents an overview of one of the most massive attempts yet made with respect to developing a model capable of making reliable predictions of the behavior of the United States socioeconomic system. One of the principal characteristics differentiating this attempt from others is that it draws upon findings from an unusually wide range of sources; the results from studies done for other purposes are integrated with additional data into a system capable of making reliable predictions with respect to a wide range of demographic and economic characteristics of the United States socioeconomic system.

The Cyert, March, and Moore paper, "A Model of Retail Ordering and Pricing by a Department Store," reports an attempt to construct an ex-

tremely detailed model of the actual decision-making process involved in ordering and pricing by a specific department in a particular large retail department store. Two questions the reader may find of use in attempting to gain some perspective of their attempt are:

1. What is the rationale underlying the study of a specific department in a specific store in immense detail, as opposed to collecting less detailed data, say, for a given department in a number of different stores?
2. Given the existence of such a model, how could the management of a specific store use it in order to learn ways of improving ordering and pricing policies?

The Kuehn and Hamburger, "A Heuristic Program for Locating Warehouses," paper outlines a heuristic approach to locating warehouses and compares it with recently published efforts at solving the problem either by means of simulation or as a variant of linear programming. The heuristic approach outlined in the paper appears to offer significant advantages in the solution of this class of problems in that it (1) provides considerable flexibility in the specification (modeling) of the problem to be solved, (2) can be used to study large-scale problems, that is, complexes with several hundred potential warehouse sites and several thousand shipment destinations, and (3) is economical with respect to computer time. The results obtained in applying the program to small-scale problems have been equal to or better than those provided by the alternative methods considered.

The Simulmatics Project*

.....

ITHIEL DE SOLA POOL
and ROBERT ABELSON†

THIS IS THE FIRST REPORT ON A PROGRAM OF RESEARCH CONDUCTED FOR the Democratic Party during the 1960 campaign. The research used a new technique for processing poll data and included computer simulation of likely voter behavior. The immediate goal of the project was to estimate rapidly, during the campaign, the probable impact upon the public, and upon small strategically important groups within the public, of different issues which might arise or which might be used by the candidates.

THE DATA

This study is a "secondary analysis" of old poll results. Students of public opinion are becoming aware that the growing backlog of earlier polls provides a powerful tool to aid in the interpretation of new poll results. Polling has now been routine for three decades, but poll archives are just beginning to be assembled. The main one is the Roper Public Opinion Research Center in Williamstown, the existence of which made feasible the project here described.¹

The first step in the project was to identify in that archive all polls anticipating the elections of 1952, 1954, 1956, and 1958. (Pre-election polls on the 1960 contest were added later when they became available.) We selected those polls which contained standard identification data on region, city size, sex, race, socio-economic status, party, and religion, the last being the item most often missing. Further, we restricted our attention to those polls which asked about vote intention and also about a substantial number of pre-selected issues such as civil rights, foreign af-

* Reprinted from the *Public Opinion Quarterly*, Vol. XXV, No. 2 (Summer, 1961), pp. 167-83.

† Massachusetts Institute of Technology and Yale University.

¹ We wish to express our gratitude to that Center, as well as to the MIT Computation Center, and to the men who originally assembled the data, especially George Gallup and Elmo Roper.

fairs, and social legislation. From 1952 to 1958 we found fifty usable surveys covering 85,000 respondents. Sixteen polls anticipating the 1960 elections were added to this number. The sixty-six surveys represented a total of well over 100,000 interviews.

PROCESSING THE DATA

To handle such massive data required substantial innovations in analytic procedures. In essence, the data were reduced to a 480-by-52 matrix. The number 480 represented voter types, each voter type being defined by socio-economic characteristics. A single voter type might be "Eastern, metropolitan, lower-income, white, Catholic, female Democrats." Another might be, "Border state, rural, upper-income, white, Protestant, male Independents." Certain types with small numbers of respondents were reconsolidated, yielding the total of 480 types actually used.

The number 52 represented what we called in our private jargon "issue clusters." Most of these were political issues, such as foreign aid, attitudes toward the United Nations, and McCarthyism. Other so-called "issue clusters" included such familiar indicators of public opinion as "Which party is better for people like you?" vote intention, and non-voting. In sum, the issue clusters were political characteristics on which the voter type would have a distribution.

One can picture the 480-by-52 matrix as containing four numbers in each cell. The first number stated the total number of persons of that voter type asked about that particular item of information. The other three numbers trichotomized those respondents into the percentages pro, anti, and undecided or confused on the issue.

We assembled such a matrix for each biennial election separately and also a consolidated matrix for all elections together. Thus, it was possible by comparison of the separate matrices to examine trends.

The reduction of the raw data to this matrix form was an arduous task. The first step was to identify in each survey those questions which seemed to bear on any of the fifty-two issue clusters we had listed as relevant to the campaign. One such cluster was attitude toward domestic communism or, as we called it for shorthand, McCarthyism. Over the past decade many questions have been asked on this and related matters in many different polls. One survey might ask, "Are you in favor of permitting a Communist to teach in the school system?" Another would ask, "What do you think of Senator McCarthy?" Another would ask, "Do you think McCarthy has done more good or harm?" The problem was to determine which questions tapped essentially the same attitude, domestic anticommunism. The decision was made by a two-step process. First, questions were grouped together *a priori* on the basis of intuitive judgment, and then this grouping was empirically tested.

The empirical test was conducted as follows: Replies to each question

were separately trichotomized. Typically, the replies had previously been coded in up to thirteen categories. Where more than three replies had been coded, the codes had to be regrouped. On the McCarthyism issue, replies were classified as McCarthyite, anti-McCarthyite, and indeterminate. A reply opposing retention of a Communist in the school system would be classified as McCarthyite. In the case of such a question as "How well do you like McCarthy?" for which a scale had originally been used, cutting points had to be set depending on the distribution.

For each pair of questions in the presumed cluster we then correlated the percentage "pro," and separately the percentage "anti," across voter types yielding two correlation matrices. (The voter types for this operation were 15, a reconsolidation of the 480. Since this operation dealt with percentages on questions from single surveys, consolidation was essential to obtain base numbers in each voter type large enough so that the percentages being correlated would be reasonably stable.) Only those questions which showed high correlations with each other were retained in a cluster. Thus, our assumption that a question about Communist teachers in the schools could be treated as equivalent to a question about McCarthy was subject to empirical validation.

In many instances questions which *a priori* seemed alike had to be discarded from the clusters. Some clusters had to be broken up into two or more. Indeed, in the particular example we have been using here, it turned out that replies to the identically worded question "How well do you like McCarthy" ceased tapping the same attitudes the minute the Senate censured him. Clusters thus represented questions which could be regarded as in some sense equivalent, both on the grounds of political common sense and on the grounds of empirical correlation.²

It should be emphasized that empirical correlation was not enough. Such a question as "Which party is better for people like you?" and a question about the image of Adlai Stevenson would correlate strongly because they were both party-linked. However, they were not included in a single issue cluster unless they also seemed politically equivalent.

The final step in the preliminary data processing—the step which gave

² We should qualify. What has been described is what we started out to do and what we did for most issue clusters. In the end, however, we were forced to compromise on certain foreign-policy clusters. This in itself is an interesting finding. On almost all domestic questions, primarily because they were party-linked or left-right linked, it was possible to validate empirically the equivalence of questions which *a priori* seemed alike. On certain foreign-policy issues this was quite impossible. The political scientist looking at a half-dozen questions about foreign aid or about the UN might conclude that they all should reflect a common underlying attitude toward that matter. However, empirically, in many instances the distribution of replies was highly sensitive to conjunctural influences or shades in wording of the question. Rather than completely abandon the hope of doing any analysis of foreign-policy issues in the campaign, we retained some clusters which failed to meet the correlational test, labeling them *a priori* clusters, not sure of what we would do with them (in fact we did very little), but feeling it better to retain them on the computer tape than to discard the data from the start.

us our matrices—was to take all cards in any one of the 480 voter types for a particular biennial period and tabulate for each issue cluster the number of replies pro, con, and indeterminate, and the number of cards on which such replies appeared. That last number varied for each cluster since some questions (e.g. turnout) were asked on virtually every survey we used, while other questions were asked only occasionally.

PURPOSES OF THE METHOD

The reader may wonder what purposes were served by reorganizing the data into the standard format just described. That handling of the data lent itself to three main uses: (1) A "data bank" was available from which one might draw the answer to any one of a vast number of questions at a moment's notice. (2) The consolidation of separate surveys made available adequate data on small, yet politically significant, sub-segments in the population. For example, we wrote a report on Northern Negro voters based upon 4,050 interviews, including 418 with middle-class Negroes. The typical national sample survey contains no more than 100 interviews with Northern Negroes, a number clearly inadequate for refined analysis. (3) The data format and its transfer to high-speed tape facilitated its use in computer simulation of the effects of hypothetical campaign strategies. This aspect of the project is the most novel and is the one to which we shall return later in this article.

THE HISTORY OF THE PROJECT

Before we illustrate those uses of the data, let us detour to examine the history of the project: the fact that it was sponsored and actually used by a partisan group makes the story of its management of some interest to students of public opinion research.

The project was initiated in the early months of 1959 by William McPhee and the authors. Our plan for computer simulation (on a different version of which McPhee had already been working)³ was presented to Mr. Edward Greenfield, a New York businessman actively engaged in Democratic politics. Through his intervention, a group of New York reform Democrats who had taken major responsibility for raising money for the Democratic Advisory Council became interested.⁴ Before this group of private individuals was willing to secure funds, however, they wanted to be sure that the results were likely to be valid and useful. In May of 1959 the project was discussed in Washington at a meeting attended by Mr. Charles Tyroler, Executive Secretary of the Democratic Advisory Council; the members of the Council executive committee; Paul Butler, Chairman of the Democratic National Committee; several

³ William McPhee, *A Model for Analyzing Macro-dynamics in Voting Systems*, Columbia University, Bureau of Applied Social Research, undated.

⁴ We wish to express our particular thanks to Thomas Finletter, Robert Benjamin, Joseph Baird, and Curtis Roosevelt for encouragement and cooperation.

other officials of that Committee; Mr. Neil Staebler, Michigan State Chairman; and a number of social science consultants, including Samuel Eldersveld, Morris Janowitz, and Robert Lane. This group was interested but reserved. It was suggested that the project should be supported for four months initially and at the end of this period a further review should be made.

The Williamstown Public Opinion Research Center agreed to permit the use of polls in their archives on two conditions: First, all basic data tabulated by Simulmatics from their cards were to be made available to the Center so the Republican Party would have an equal opportunity to use such data if they wanted them. We provided a print-out of the data on the computer tape, but not, of course, the programs for simulation nor supplementary data obtained from other sources (e.g. the census) and used in our system. Second, and demanded by both the Roper Public Opinion Research Center and the social scientists engaged in the study, all results could be published for scientific purposes after the election. This article is part of our program to meet that condition.

Given the green light to carry out the project, the principals organized themselves as The Simulmatics Corporation, for although the objective of the project constituted scientific research, it was clear that universities would not and should not accept financing from politically motivated sources or permit a university project to play an active role in supplying campaign advice to one party.

The summer of 1959 was devoted to the data reduction job described above. In October 1959, when the preliminary data processing had been substantially completed, a review meeting in New York was attended by many of the same persons who had been at the Advisory Council meeting in May plus a number of social science consultants, including Harold Lasswell, Paul Lazarsfeld, Morris Janowitz, and John Tukey. Although the degree of confidence in the basic approach ranged from enthusiasm to doubt, a decision to proceed was quickly reached.

The next step was the development of computer programs, some of which will be discussed below. One objective was to make possible rapid incorporation of new data which might, we hoped, become available during the campaign. Our hope, as we shall see, was only slightly fulfilled.

By June of 1960 we were able to prepare a first report as a sample of the kind of thing which might be done by the Simulmatics process. That was the report on the Negro vote in the North.

Our contractual arrangements with our sponsors ended with the preparation of the process and of this report illustrating it, shortly before the 1960 convention. It was understood that actual use of the service in the form of further reports on specific topics would be purchased by appropriate elements of the party in the pre-campaign and campaign period at their discretion. In the immediate pre-convention period, the National Committee felt that it should not make decisions which would shortly be

the business of the nominee. After the convention, the Kennedy organization, contrary to the image created by the press, did not enter the campaign as a well-oiled machine with a well-planned strategy. Except for the registration drive, which had been carefully prepared by Lawrence O'Brien, no strategic or organizational plan existed the day after the nomination. It took until August for the organization to shake down. No campaign research of any significant sort was therefore done in the two months from mid-June to mid-August, either by Simulmatics or by others. In August, a decision was made to ask Louis Harris to make thirty state surveys for the Kennedy campaign. However, because of the late start, data from these surveys would not be available until after Labor Day. On August 11, the National Committee asked The Simulmatics Corporation to prepare three reports: one each on the image of Kennedy, the image of Nixon, and foreign policy as a campaign issue. These three reports were to be delivered in two weeks for use in campaign planning. Along with them we were to conduct a national sample survey which, in the minds of the political decision makers, would serve to bring the Simulmatics data, based as they were on old polls, up to date. (It should be mentioned that one of the most difficult tasks of the Simulmatics project was persuading campaign strategists that data other than current intelligence could be useful to them.) The national survey by telephone was conducted for the project by the Furst Survey Research Center and was indeed extremely useful in guiding the use of the older data. It confirmed the published Gallup finding that Nixon was at that point well in the lead, though we disagreed on the proportion of undecideds (we found 23 percent). It made us aware that Nixon's lead was due to women. It also persuaded us that voters were largely focusing upon foreign policy at that point in the campaign.

The relationship between the use of such current intelligence and the use of a simulation model developed out of historical data is analogous to the relationship between a climatological model and current weather information. One can predict tomorrow's weather best if one has both historical information about patterns and current information about where one stands in a pattern. While it would be presumptuous to assert that in two weeks of intense activity we approached an effective integration of the two sets of data, that was the ideal we had in mind and which in some limited respects we approximated.

It should be added that the introduction of the national survey data was possible only because of prior preparation for rapid data analysis. The survey was ordered on a Thursday, the field interviewing took place between Saturday and the following Thursday, by Friday morning all cards had been punched, and by Friday night the pre-programmed analysis had been run and preliminary results were given to the National Committee.

The three reports that had been ordered on August 11 were delivered

on August 25. The speed of the entire operation is, of course, a testimony to the advantages of a high-speed computer system. Nonetheless, such intense pressure is not an optimum condition for research work, even though rapid analysis was one of our objectives from the start. The reader who suspects that under those circumstances clerical errors inevitably occurred is quite right. It was our good fortune that none of those which we have found since in rechecking have turned out to alter any conclusion, but we do not recommend such limited schedules as a normal mode of work. Nevertheless, with well-prepared computerized analysis, it can be done when necessary.

The reader may ask whether the large preparatory investment was justified in terms of the quantitatively limited use of the project. When we planned the project, we—perhaps unrealistically—anticipated active campaign work from the beginning of the summer until about September 15. (Anything done later than that would hardly be useful.) How far the investment was justified by the two weeks of work actually done is a question which we find impossible to answer. An answer depends on an estimate of how much impact the contents of the reports had on the campaign. The reports received an extremely limited elite circulation. They were seen during the campaign by perhaps a dozen to fifteen key decision makers, but they were read intelligently by these talented and literate men.

Despite the contraction of our effort, our own feeling is one of relative satisfaction that the Simulmatics project was able to provide research on demand concerning the key issues at perhaps the critical moment of the campaign. While campaign strategy, except on a few points, conformed rather closely to the advice in the more than one hundred pages of the three reports, we know full well that this was by no means because of the reports. Others besides ourselves had similar ideas. Yet, if the reports strengthened right decisions on a few critical items, we would consider the investment justified.

EXAMPLES OF USE OF THE SYSTEM

Earlier in this article we listed three uses of the method herein described: providing a "data bank," rapidly available; providing data on small, politically significant groups; permitting computer simulation. The first of these advantages has perhaps already been adequately illustrated. Let us turn to the other two.

Our report on Northern Negro voters did not use a computer simulation but rather illustrated the capability of the process to provide information about small subgroups of the population. Compare here a number of quotations from the report with what we could have said working from a single survey containing responses from perhaps 100 Northern Negroes. The report demonstrated, for example, that between 1954 and 1956

[A] small but significant shift to the Republicans occurred among Northern Negroes, which cost the Democrats about 1 percent of the total votes in 8 key states [a shift which continued in 1958]. In those years, the Democratic Party loss to the Republican Party was about 7 percent of the Northern Negro vote—enough to cause a one half percent loss in the *total* popular vote in the eight key states. In addition, among Northern Negro Independents, only about one quarter actually voted Republican in 1952, but about half voted Republican in 1956, enough of a shift to cause an additional loss of a little less than one half percent of the total popular vote in the eight key states.

.

The shift against the Democrats is more marked among the opinion leading middle class Negroes than among lower-income Negroes.

.

Anti-catholicism is less prevalent among Negroes than among Northern, urban, Protestant whites.

.

The most significant point of all is the fact that the shift is not an Ike-shift: it is a Republican Party shift. It affects Congressional votes as much as Presidential votes.

In addition, the report demonstrated that Northern urban Negroes vote as often as whites of comparable socio-economic status, and that "there is no sharp difference between Negroes and comparable whites in their feelings about Nixon."

This report was made available to all the leading Democratic candidates, to the Democratic National Committee, and to the drafters of the Democratic platform. Probably no one can say what influence, if any, it had upon them. Those men themselves would not know which of the many things they read or heard shaped their decisions. As outside observers, we can assert only that the report was placed in the hands of the platform framers in the ten days preceding the drafting of the problem, and was read.

The most dramatic result, however, was, as indicated above, the finding that Eisenhower had not generated among Negroes the kind of personal following that he had among most white voter types. This suggested that the Negro vote presented far more of a problem to the Democratic campaign than appeared at first glance; it could not be assumed that the losses in recent years would be recovered with Eisenhower out of the picture.

SIMULATIONS

We turn now to what was perhaps the most novel aspect of the study—the use of computer simulations. We describe, first, how we simulated state-by-state results and, second, how we simulated the impact of the religious issue.

One of the benefits gained from the large number of interviews we used was the possibility of approximating state-by-state results. A national sample survey—even a relatively large one—has too few cases from most states to permit any significant analysis of state politics. The same would have been true, however, even for our voluminous data if we had attempted to do a state-by-state analysis in a simple way. We had an average of about 2,000 interviews per state, but that is a misleading figure. In a small state there might have been no more than 300 or 400 interviews, and on a particular issue cluster that had occurred, for example, in only one-tenth of the surveys, there would be too few cases for effective analysis. We therefore developed a system for creating synthetic, or simulated, states.

By an elaborate analysis of census, poll, and voting data—made more difficult because 1960 census results were not yet available—we developed a set of estimates on the number of persons of each voter type in each state. (Note that since *region* was one of the defining characteristics for the 480 voter types, there were at most only 108 voter types in any given state.) It was assumed that a voter of a given voter type would be identical regardless of the state from which he came. A simulated state therefore consisted of a weighted average of the behaviors of the voter types in that state, the weighting being proportional to the numbers of such persons in that state. For example, we thus assumed that the difference between Maine and New York is not truly a difference between New Yorkers and inhabitants of Maine as such, but a difference in the proportions of different voter types which make up each state. We assumed that an “upper-income Protestant Republican rural white male” was the same in either state, and that a “small-city Catholic Democratic lower-income female” was also the same in either. This assumption enabled us to use all cases of a voter type from a particular region in arriving at a conclusion for a state.

We do not assert that the assumptions on which this simulation is based are true. On the contrary, we can be sure that they are partly false. The interesting question intellectually is how good were the results obtained with these partially true assumptions. The test is, of course, how far state-by-state predictions made on these assumptions turn out to correspond to reality. To the extent that they do, they suggest that the essential differences between states in a region are in distributions of types rather than in geographic differences, even within a voter type.⁵

Upon this simulation of states was built a second and more interesting simulation, one which attempted to assess the impact of the religious is-

⁵ The states where the simulation was most notably off included Arizona, Nevada, New Mexico, Idaho, and Colorado, states mostly of small population, and states which, in the absence of a “Mountain Region” in our classification, we attempted to treat as Western or Midwestern. Clearly, the assumption of regional uniformity was misleading as applied to them.

sue. Since the one simulation rests upon the other, the effectiveness of the state simulation is simultaneously tested by examination of the religious simulation. The latter, the main simulation actually carried out during the campaign, represented a hypothetical campaign in which the only issues were party and Catholicism. Our report of this simulation was limited to the North because of the peculiar role of party in the South. The outcome was a ranking of thirty-two states ranging from the one in which we estimated Kennedy would do best to the one in which we estimated he would do worst. The ranking was:

- | | |
|------------------|-------------------|
| 1. Rhode Island | 17. Pennsylvania |
| 2. Massachusetts | 18. Nevada |
| 3. New Mexico | 19. Washington |
| 4. Connecticut | 20. New Hampshire |
| 5. New York | 21. Wyoming |
| 6. Illinois | 22. Oregon |
| 7. New Jersey | 23. North Dakota |
| 8. California | 24. Nebraska |
| 9. Arizona | 25. Indiana |
| 10. Michigan | 26. South Dakota |
| 11. Wisconsin | 27. Vermont |
| 12. Colorado | 28. Iowa |
| 13. Ohio | 29. Kansas |
| 14. Montana | 30. Utah |
| 15. Minnesota | 31. Idaho |
| 16. Missouri | 32. Maine |

The product-moment correlation over states between the Kennedy index on the simulation (not strictly speaking a percent) and the actual Kennedy vote in the election was .82. It should be emphasized that this satisfying result was based upon political data not a single item of which was later than October 1958. Surveys on the 1960 election were not available soon enough to be incorporated into this analysis.

The basic method in this simulation was a fairly straightforward application of the cross-pressure findings of earlier election studies.⁶ These findings enabled us to improve our estimate of how a particular voter will behave if we know the cross-pressures he is under. With such knowledge, an analyst should feel more comfortable making guesses about how voters under particular kinds of cross-pressure will shift in an election than he would about making an over-all intuitive guess at the outcome. The method of this simulation was to make a series of such detailed estimates and then let the computer put them together to give an over-all outcome.

To make these detailed estimates we classified our set of 480 voter types into 9 possible cross-pressure subsets arising from a 3-by-3 breakdown on religion and party: Protestants, Catholics, and others; Republicans, Democrats, and Independents. For each of the nine resulting situa-

⁶ Bernard R. Berelson, Paul F. Lazarsfeld, William N. McPhee, *Voting: A Study of Opinion Formation in a Presidential Campaign* (Chicago: University of Chicago Press, 1954).

tions we made a prediction. For example, take the Protestant Republicans. They were not under cross-pressure. Since our data had revealed no substantial dislike of Nixon as an individual among such voters, we saw no reason why their vote in 1960 should differ substantially from their vote in 1956, even though Eisenhower was not running. Thus for them we wrote two equations:

$$\begin{aligned} V_k &= P_{56}(1 - P_{35}) \\ V_n &= Q_{56}(1 - P_{35}) \end{aligned}$$

meaning that the predicted Kennedy percentage (V_k) in any voter type of this Protestant-Republican sort would be the percentage of persons in that voter type who had indicated a preference for Stevenson in the 1956 polls (P_{56}), reduced by the nonvoting record of that voter type ($1 - P_{35}$).⁷ The equation for the expected Nixon percentage (V_n) was the same except that it used the 1956 Eisenhower supporters (Q_{56}).

The above was the simplest set of equations used. Let us now turn to a more complicated set, that for a group under cross-pressure—Protestant Democrats. First, we decided that, barring the religious issue, 1958 vote intentions would be a better index of the Protestant Democrats' 1960 vote than would their 1956 vote intentions. Too many of them were Eisenhower defectors in 1956 for us to believe that 1956 was a good indicator of normal behavior. On the other hand, 1958 polls would overestimate their Democratic vote, since many of them would defect again against a Catholic. However, it would not suffice merely to subtract the percentage who gave anti-Catholic replies on poll questions, for perhaps those very Democrats who were anti-Catholic were the ones who in practice voted Republican anyway. In short, the question was: Were the bigot defectors right wingers whose vote the Democrats would lose even without a Catholic candidate? Our system could not give us that information for each respondent incorporated into our data. While one respondent in a voter type might have been polled in a survey in 1958 about his vote intentions, another man of the same voter type, on a different survey, might have been polled on whether he would vote for a Catholic for President. To estimate the correlation between these two variables we had to find one or more surveys on which both questions appeared. We then ran anti-Catholicism by 1958 vote for each of the more numerous Protestant Democrat voter types. We found that among them the ratio ad/bc in the following fourfold table averaged about .6. With that information we could estimate how many of the anti-Catholics were hopeless cases anyhow (i.e. had gone Republican even in 1958) and how many would be net losses only in a campaign dominated by the religious issue.

⁷ Since we trichotomized results, $P_{56} + Q_{56}$ do not add up to 100 percent. The reader may wonder why a turnout correction is added: are not the residuals the nonvoters? The answer is that a turnout correction is needed because many more persons express a candidate preference on a poll than actually turn out to vote.

1958 Vote Intentions	Anti- Catholic	Not Anti- Catholic
Democratic.....	<i>a</i>	<i>b</i>
Republican.....	<i>c</i>	<i>d</i>

It should be added here that we decided to take poll replies on the religious issue at face value. We were not so naïve as to believe that this was realistic, but since we were not trying to predict absolute percentages, but only relative ones, all that mattered was that the true extent of anti-Catholicism, voter type by voter type, should be linearly related to the percentage overtly expressed. Even this could only be assumed as a promising guess.

Finally, in predicting the vote of the Protestant Democrat voter types, we took account of the established finding that voters under cross-pressure stay home on election day more often than voters whose pressures are consistent. Therefore, for our 1960 estimate we doubled the historically established nonvoting index for these types.

Thus we arrived at equations applied to each Protestant Democratic voter type:

$$V_k = (P_{58} - a)(1 - 2P_{35})$$

$$V_n = (Q_{58} + a)(1 - 2P_{35})$$

The estimate of anti-Catholic 1958 Democratic voters (i.e. persons in cell *a* in the fourfold table above) was arrived at by the computer, given that

$$a + b = P_{58} \quad a + c = P_{14}(P_{58} + Q_{58}) \quad \text{and } \frac{ad}{bc} = .6$$

$$P_{14} = \text{percent anti-Catholic}$$

Space precludes a similar examination of each of the other of the nine conditions.⁸ Suffice it to say that one other set of serious guesses had to be made, namely what proportion of those Democratic Catholics who had voted Republican in 1958 would switch back to their party to vote for Kennedy and what proportion of Republican Catholics who had voted Republican in 1958 would also switch to Kennedy. After an exami-

⁸ With the above information, the remaining equations should be decipherable and are reported here for the record:

Protestant Independents, same equations as Protestant Democrats.
Catholic Democrats and Catholic Independents:

$$V_k = \left(P_{58} + \frac{Q_{58}}{3} \right) (1 - P_{35})$$

$$V_n = \frac{2Q_{58}}{3} (1 - P_{35})$$

Catholic Republicans:

$$V_k = \left(P_{58} + \frac{Q_{58}}{3} \right) (1 - 2P_{35})$$

$$V_n = \frac{2Q_{58}}{3} (1 - 3P_{35})$$

(cont.)

nation of the trial-heat data from polls which asked about Kennedy vs. Nixon, we decided to use one-third as the proportion in each case, and to use that figure also as an estimate of the proportion of Catholic independents who would be won back by the religious issue.

The simulation required that the computer make 480 separate calculations, each one using the appropriate set of equations from above. During each of the 480 calculations, the computer put into the equations values for turnout record, 1958 vote intention, 1956 vote intention, and anti-Catholicism, derived from the data which had been assembled about that particular voter type. This gave a 1960 vote estimate for each voter type for the particular hypothetical campaign being investigated. Weighted averages of these gave the state-by-state estimates.

These estimates, as we have already noted, turned out to be close to the actual November outcome. They were not intended to be predictions. Or, rather, they were *contingent* predictions only. They were predictions of what would happen *if* the religious issue dominated the campaign. We did not predict that this would happen. We were describing one out of a set of possible types of campaign situation. But by August, when we took our national survey, comparison of our simulation and the survey results showed that this situation was actually beginning to occur. And the closeness of our contingent prediction to the final November result suggests that, indeed, the religious issue was of prime importance.

How close was the religious-issue simulation to the actual outcome compared to alternative bases of prediction? A full exploration of this remains to be made. We must, for example, further vary the parameters used in the simulation to determine which ones affect the results most critically and which values of those give the best prediction. For the present we look only at the one set of values and equations on which we relied during the campaign and which has already been described. (A few variations were tried and dismissed during the campaign, but none that made much difference.) How did this one simulation compare with other predictive data?

All others:

$$\frac{ad}{bc} = .6$$

$$a + b = \frac{P_{58} + P_{56}}{2}$$

$$c + d = \frac{Q_{58} + Q_{56}}{2}$$

$$a + c = P_{14} \frac{(P_{58} + Q_{58} + P_{56} + Q_{56})}{2}$$

$$b + d = (1 - P_{14}) \frac{(P_{58} + Q_{58} + P_{56} + Q_{56})}{2}$$

$$V_k = \left(\frac{P_{58} + P_{56}}{2} - a \right) \left((1 - 2P_{36}) \right)$$

$$V_n = \left(\frac{Q_{58} + Q_{56}}{2} + a \right) \left((1 - 2P_{36}) \right)$$

An obvious comparison is with the Kennedy-Nixon trial heats on polls taken at the same time as the latest polls used in the simulation. The correlation between the state-by-state result of these polls and the actual outcome is but .53 as compared to .82 for the simulation. The simulation, in short, portrayed trends which actually took place between the time the data were collected and election day. The uncorrected polls two years before the election explained but one-fourth of the variance in the real results, while intelligent use of them taking into account the cross-pressure theory of voting behavior allowed us to explain nearly two-thirds of the variance.

A more stringent comparison would be with Kennedy-Nixon trial heats run in August 1960, when the simulation was run on the computer. Such a comparison would answer the question of whether the Democratic Party would have gotten as good information at that date by the conventional means of up-to-the-minute field interviewing as it got by re-analysis of old data. Very likely it could have, if it had chosen to invest in a large enough national sample survey to give it state-by-state results, for as far as we can now tell the Catholic issue exerted most of its impact by shortly after the conventions. However, until poll data for that period becomes available we can only speculate. We wish to emphasize, however, that at some point in the history of the campaign, poll data certainly came into close correlation with the November election results and thus with our simulation. The date the raw poll results became as or more predictive than the simulation would be the point in the campaign at which mechanisms of voter behavior anticipated in the simulation became reality.

Besides simulation and polls, what other indices might have forecast long in advance the state-by-state order of voting in 1960? Results of previous elections would be one such index. Perhaps the rank order of the states in a previous election is a good forecast of rank order in future ones, even if the electoral outcome changes. (The whole country could move one way or the other, leaving the order of the states much the same.) But, if one is to use this device, which election should one use? The year 1956 was a presidential election year, as was 1960, but in 1956 the Eisenhower phenomenon was operating. 1958, although more recent and less affected by Eisenhower's idiosyncratic appeal, was a Congressional election year. In our simulation, too, we faced this problem. We resolved it for some voter types one way, for some another. But what happens if one relies on a single simple over-all assumption of continuity between elections? The result is not very good, though slightly better using 1956 than 1958. The product-moment correlation of Northern results between 1956 and 1960 was .39, between 1958 and 1960, .37. The multiple correlation using both earlier years was .44 with the 1960 election. So far our simulation clearly was superior as a forecast.

Perhaps one might have made a good prediction of the impact of the

religious issue by a simple slide-rule method of calculation instead of by an elaborate computer procedure. One could correct the 1956 or 1958 vote by some crude percentage of the Catholic population of each state. That would have worked and worked well if, by some act of intuitive insight, one could have hit on the right percentage correction. One would have had to decide first of all to use 1956, not 1958, as the base, for no simple correction of the 1958 results gives a good correlation with the actual outcome. If one had that correct flash of intuition, one could have surpassed our complex simulation with a correction of exactly 34 per cent of the Catholic percentage of the population added to the Democratic vote. The correlation with the actual outcome achieved by this process is .83. The simulation was better, however, than any correction except 34 per cent. It is better than 33 or 35 per cent. At corrections of 32 and 40 per cent the coefficients of correlation for the simple correction procedure drop below .80.

There was, in other words, a "lucky guess" way of estimating the effect of the religious issue in the campaign which would have given an excellent prediction. But even if we had tried to make such an over-all estimate and had somehow arrived at the right "lucky guess," we could not have defended it against skeptics. What the simulation did was to allow competent political analysts, operating without inspired guesses, to make sober, scientifically explicable estimates that they were willing to commit to paper before the facts. As the accompanying table shows, the simulation gave results about as good as the very best which *hindsight* now tells us could have been reached by simpler methods if infused by the right lucky guesses.

Correlations with Actual Election Results

Trial heats contemporaneous with simulation data.....	0.53
Continuity with 1956.....	0.39
Continuity with 1958.....	0.37
Continuity with 1956 and 1958.....	0.44
1956 results with optimum, or "lucky guess," correction for Catholic vote.....	0.83
Simulation as done during campaign.....	0.82

The essence of the simulation was to treat each voter type separately. Under what conditions should one expect that procedure to obtain a better result than an optimal across-the-board correction applied to the total? Clearly, if the process at work in each voter type was uniform it would make no difference whether we applied correction factors voter type by voter type or to the total. One could add 34 per cent of each Catholic voter type to the Democratic vote for that type or add 34 per cent of total Catholics to the total Democratic vote and come out with the same result. Where there are complex interactions of several variables on a voter type, however, then calculations done the two ways are no longer equal. If, for example, turnout varies between voter types and party vot-

ing also varies, then an equation applied to each voter type could not equally well be applied to total voters.

It is clear that we did not use the most predictive values for all parameters in our simulation. Determining what these were with the aid of hindsight is part of our present research program. But before election day we had no way of knowing what they were. (The one-third of 1958 Catholics casting Republican votes likely to go Democratic in 1960 according to our equations should not be confused with the 34 per cent of all 1956 Catholic voters, which turned out to be a good across-the-board correction.) The fact that our result came out on a par with the optimum simple correction which hindsight has enabled us to make is a crude measure of the gain from working voter type by voter type, with account taken of interactions within each type, that is, the gain from the computer operations.

The test of any new method of research is successful use. The outcome of the present study gives reason to hope that computer simulation may indeed open up the possibility of using survey data in ways far more complex than has been customary in the past. The political "pros" who commissioned this abstruse study were daring men to gamble on the use of a new and untried technique in the heat of a campaign. The researchers who undertook this job faced a rigorous test, for they undertook to do both basic and applied research at once. The study relied upon social science theories and data to represent the complexity of actual human behavior to a degree that would permit the explicit presentation of the consequences of policy alternatives.

This kind of research could not have been conducted ten years ago. Three new elements have entered the picture to make it possible: first, a body of sociological and psychological theories about voting and other decisions; second, a vast mine of empirical survey data now for the first time available in an archive; third, the existence of high-speed computers with large memories. The social science theories allow us to specify with some confidence what processes will come to work in a decision situation. The backlog of survey data permits us to estimate the parameters of these processes with fair precision and great detail for each small element of our national population. The computer makes possible the handling of this mine of data. More important still, it makes possible the precise carrying out of long and complex chains of reasoning about the interactions among the different processes. In summary, we believe that conditions now exist for use of survey data in research far more ambitious than social scientists are used to. If it is possible to reproduce, through computer simulation, much of the complexity of a whole society going through processes of change, and to do so rapidly, then the opportunities to put social science to work are vastly increased. It is our belief that this is now possible which was put to a test by the campaign research reported here.

A Stochastic Microanalytic Model of a Socioeconomic System*

.....

GUY H. ORCUTT, MARTIN GREENBERGER, JOHN KORBEL,
and ALICE M. RIVLIN†

There is reason to hope that the reliability of predictions being made with respect to the United States socioeconomic system might be improved if they were generated by a model constructed in terms of the behavior and interaction of the fundamental units of that system. This paper, which describes such a model of the U.S. socioeconomic system, sets forth its essential formal aspects and indicates the correspondence of these aspects with the system being modeled. One of these fundamental formal aspects is the fact that the model involves components, variables relating to the components, and relationships between these variables. The actual names given to the components and to the variables are of minor importance except insofar as they serve to indicate the correspondence between features of the model and features of the real economy.

A distinctive characteristic of the type of model described in this paper is that it contains components corresponding to microcomponents of the real socioeconomic system. The components in the model may be classified into the following three broad categories: decision-making units; markets; and things which are consumed, held, used in productive processes, or which enter into transactions.

In the real system we can identify several different types of decision-making units—individuals, families, firms, banks, labor unions, local governments. The model, like the real system, contains a population of decision-making units composed of a relatively small number of different types of such units and a relatively large number of units of each type.

* Reprinted with minor modifications from Chapter 2 of Guy H. Orcutt, Martin Greenberger, John Korbel, and Alice M. Rivlin, *Microanalysis of Socioeconomic Systems: A Simulation Study* (New York: Harper & Bros., 1961), pp. 13-41.

† University of Wisconsin, Massachusetts Institute of Technology, University of Wisconsin, and Washington, D.C.

In the real system we can identify many different markets. The model, as does the real system, contains many different markets, and these components provide linkages between the decision-making units in the model just as real markets serve to link potential buyers and sellers or borrowers and lenders in the real system.

In the real system we can identify the many different types of goods, services, credit instruments, and other things which are used or held or which enter into transactions. The model, like the real system, contains a constantly evolving population of these things which are used or held or produced by decision-making units and which may pass through markets on their way between decision-making units.

Variables in the model relate in one way or another to the components of the model, and the name of a variable should be explicit enough to make clear the component or components to which it relates. It is convenient to classify variables into the following three broad categories: output variables, input variables, and status variables.

The decision units have various possible behavior outputs. Individuals may enter the labor force, earn wages, get married; married couples may set up new households, have children, get divorced; firms may purchase raw materials, hire labor, establish new plants, produce goods; and so forth. Any observable behavior of a decision-making unit—even the expression of an opinion or attitude—could be considered an output.

The outputs of a decision-making unit during a given time period are taken to depend on prior inputs to the component and on the values of the component's status variables as of the beginning of the period.

Anything external to a decision-making unit which acts on it or influences its behavior may be considered an input to the component. Inputs may thus include the seasons, the weather, time, and the prior outputs of other components.

The component's status variables are internal variables which describe the state of the component. The values of these variables as of the beginning of a time period may influence the behavior of the decision-making unit during the time period, and the output of the decision-making unit during the period may alter the values of the status variables as of the start of the following period. Status variables are characteristics of the units themselves—initially assigned in the proportions in which these characteristics appear in the real system. For example, status variables of individuals might include age, sex, marital status, education, income; status variables of firms might include inventories, sales in the last quarter, back orders, balance-sheet variables, even anticipated sales.

In addition to components and to the variables relating to the components, a model of an economy must also contain relationships if it is to generate any predictions. Relationships specify how the values of different variables in the model are related to each other or how they are

otherwise generated. Relationships are of two broad types: identities and operating characteristics.

Identities are accounting or tautological statements that may be introduced for convenience. Thus total sales of a product are set equal to total purchases, total assets of a firm are set equal to total liabilities and equity.

An operating characteristic is a relationship specific to a given component which specifies either an hypothesis or an assumption about how an output variable of the component is related to status and input variables of the component. Operating characteristics, by specifying how components respond to stimuli or how they generate outputs even in the absence of external stimuli, embody most of the real knowledge brought to bear, and it is primarily upon them that any predictive use of the model must ultimately depend. And whereas components and variables relating to components may be directly observed in the real system, it is important to recognize that operating characteristics cannot be directly observed but rather must be inferred by inductive methods of analysis. Operating characteristics may be of any form research indicates to be appropriate.

In general the relations linking outputs of a decision-making unit to the prior inputs to the unit and to the status variables of the unit are considered to be stochastic in nature; that is, it is the probabilities of occurrence of certain outputs, rather than the outputs themselves, which are regarded as functions of the input and the status variables. The probability functions which relate outputs to input and status variables are called the operating characteristics of the component. If the probability that an individual will marry in a given period is taken to depend on his age, sex, and marital status, then the table (or function) which specifies the probabilities of marriage for males and females of various ages and marital conditions would be one of the operating characteristics of individuals.

This probabilistic approach to predicting behavior does not reflect a philosophical position about the nature of causation. It simply represents an attempt to utilize the kind of information which we actually have about the decision-making units of our economic system. To take an obvious example, consider the amount of information one would need to predict whether or not a particular individual will be alive a year from now. One would need his complete medical history and that of everyone with whom he came in contact, all his actions in the period and those of others, and so forth. Clearly the problem is impossible. Nevertheless, when large populations are "at risk," the vital events exhibit considerable regularity. An experienced life insurance actuary can predict within narrow limits and with a very high degree of confidence the number of his company's insurees who will die in the next twelve months. To do this he needs only a modest amount of information about the insured

population—perhaps only its distribution by age and sex. He estimates the probability of death for individuals of given age and sex from past experience and assumes that death will continue to occur according to these estimated probabilities.

An actuarial approach also seems appropriate to predicting many types of economic behavior. To predict whether a particular household will or will not buy a house in the next month, for example, one would need to know the complete preference scales of the members, the “power structure” within the household, and so forth. Even if one could measure these things, it would not be feasible to do so for very many households. It may be feasible, however, to estimate probabilities of house purchase for groups of households with certain observable characteristics. It is this type of information that the model utilizes.

One advantage in describing a model in terms of components, variables relating to the components, and relationships between the variables is that it facilitates a building-block approach to the construction and testing out of the model. The components serve rather naturally as the major building blocks and facilitate a useful subdivision of the overall effort required to achieve a useful model of the entire system. The building of models of each kind of component may then proceed with only a very limited coordination required between the groups working on different components. Each research group needs to be aware of what outputs are being generated by other components and what outputs are expected from their kind of component, but that is all. They are free to be guided solely by the evidence their research turns up in generating the outputs expected from their components. Research coordination at a systems level plays an essential role only in those situations where researchers on a component find it essential to use, as inputs, variables which are not being assumed as given or as being generated by other components in the system.

We now present a hypothetical model of a bank—but it cannot be emphasized too strongly that we do so solely to illustrate the use of terms and to show some general features that might be found in a model of any component.

HYPOTHETICAL MODEL OF A BANK

I. Variables

All liabilities of the bank are treated as deposit liabilities; all non-earning assets are treated as reserves; and all earning assets are treated as loans. Notice that there are 11 variables of which 5 are exogenous insofar as the model of this bank is concerned.

A. Input

1. *CHK_t* *CHecKs* written on bank by depositors during time period *t*
2. *MDP_t* *Money DePosited* during time period *t*

3. DTI_t DebT to bank Incurred during time period t
4. MRP_t Money RePaid to bank during time period t
5. RRR_t Required Reserve Ratio during time period t

B. Balance Sheet (Status Variables)

1. DP_t DePosit liabilities of bank at start of time period t
2. RS_t ReServe assets of bank at start of time period t

C. Outputs

1. MPO_t Money Paid Out of bank at request of depositors during time period t
2. MLN_t Money LoaNed by bank during time period t
3. DTE_t DebT to bank Extinguished by bank during time period t
4. INR_t INterest Rates required by bank on debt incurred during time period t

II. Relationships

A. Identities

1. $DP_{t+1} \equiv DP_t + MDP_t - CHK_t$
2. $RS_{t+1} \equiv RS_t - MPO_t - MLN_t + MDP_t + MRP_t$

B. Operating Characteristics

1. $INR_t = F_1(RS_t, DP_t, RRR_t)$

In words, this first operating characteristic stated that the management of this bank behaves in such a fashion that the interest rate required by this bank, on debt incurred during time period t , is some function of the reserve assets of the bank at the start of time period t , the deposit liabilities of the bank at the start of time period t , and the required reserve ratio for this bank during time period t .

2. $MPO_t = CHK_t$
3. $MLN_t = F_2(DTI_t)$
4. $DTE_t = MRP_t$

The above specification of a model of a bank is incomplete in that the forms of functions 1 and 2 have not been given. As a matter of fact, they only could be specified on the basis of extensive research regarding the behavior of banks. And such research would probably show that additional input variables and/or status variables are required to account adequately for the way in which a bank sets the interest rate it requires on debt incurred to it during any given time period. It also is obvious that, in setting the terms and conditions under which it is prepared to loan, a bank does much more than merely set one or more interest rates.

In the foregoing bank model, the managers can affect only indirectly their bank's investment in earning assets. To do this they must influence the behavior of potential borrowers from the bank by altering the terms under which loans are extended. In a more realistic model, earning assets would include a variety of assets which could be sold or bought in the open market. This would permit managers to make direct adjustments

of their portfolio positions and would call for substantial alteration of the model with respect to output variables generated and input variables used. It also is clear that in a more realistic model information about alternative earning possibilities would be a necessary input.

For all of its incompleteness and unreality, certain requirements in building a model of a component should be apparent from an examination of the above model. In the first place it is essential to distinguish clearly those things the component is actually assumed to be doing and those things to which it is assumed to be responding. The things that the component does are described in terms of the output variables of the component. The things to which it responds are described in terms of input variables and status variables of the component. The input variables are to be taken as given insofar as this particular component is concerned, although of course either they will be determined by the outputs of other components in the system or they will be variables which are treated as exogenous to the whole system. The status variables, on the other hand, are internal to this particular component, and if they are required as of the beginning of each period, then the operating characteristics of the component must serve to update them each time period as well as to generate the output variables. Values of input variables and status variables must always be available prior to the generation of values of associated output variables.

As an aid in seeing how a model of a component generates output variables and updates status variables, it is often useful to represent certain features of the model by means of a flow diagram. A flow diagram corresponding to our model of a bank is shown in Figure 1. Lines with arrowheads are used to show where variables come from and where their influences operate. In the case of the status variables these lines have been omitted in order to avoid unduly complicating the diagram. If one line splits into two lines at a heavy dot, the variable whose influence is indicated through the single line is to be thought of as operating through each of the two new lines. If a line crosses another line with a little semi-circular movement, then no connection of any sort is implied.

For purposes of graphically representing the entire system it is desirable to use a vastly simplified representation of a decision-making unit. In general, decision-making units generate many output variables and these output variables are transmitted and distributed to other decision-making units by components called markets. Leaving out all internal detail, a decision-making unit such as a bank or a family unit might be represented by Figure 2.

It is intended that Figure 2 convey the impression of a slab or block representing a decision-making unit which generates an unspecified number of output variables relevant to each of the M markets. Those output variables of this decision-making unit that are labeled *to Mkt. 1* will of course show up as inputs into market 1. With respect to any spe-

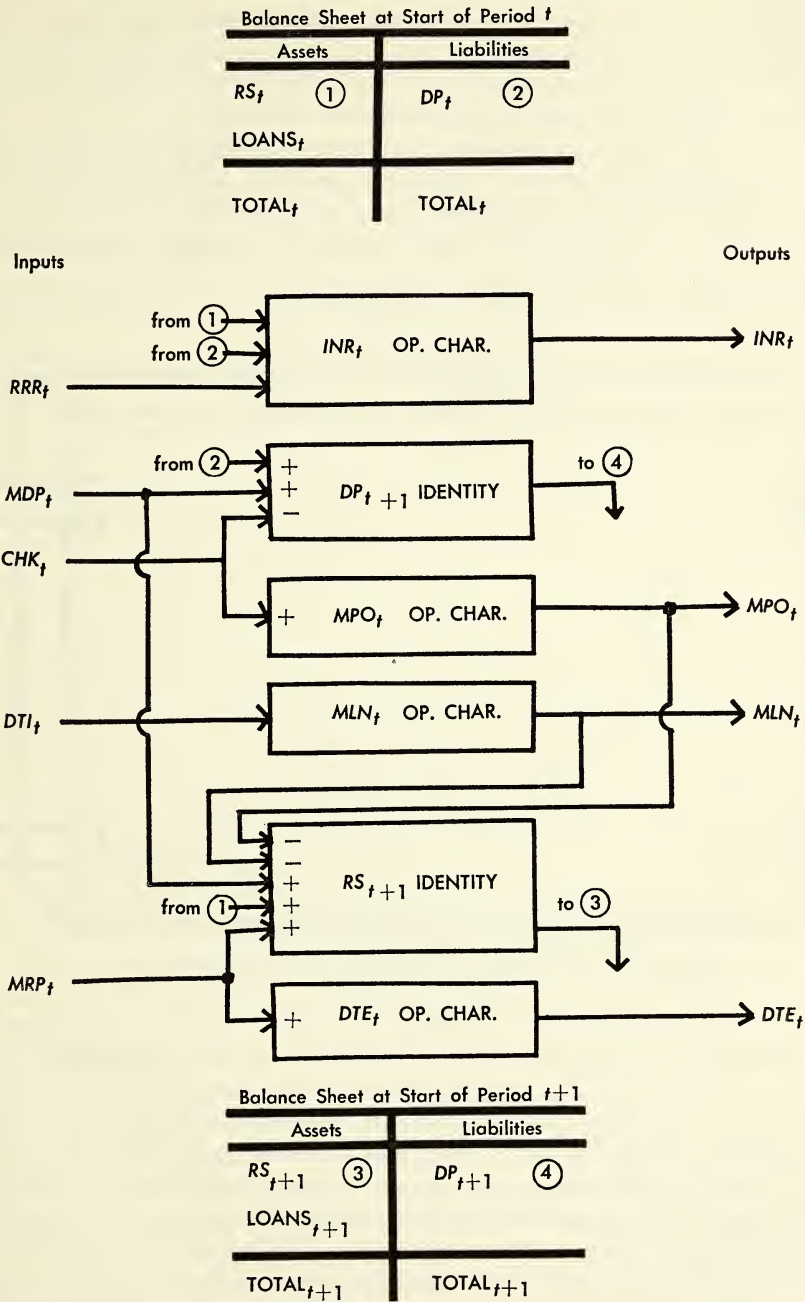


FIGURE 1. Flow diagram of a model of a bank.

cific market, there may be several such variables operating from any particular unit or there may not be any. In a similar way the input variables of this decision-making unit are to be thought of in general as having originated as output variables of some market.

It is now necessary to elucidate the matter of input-output linkage in the model. Some outputs are simply end products and nothing more is done with them except to aggregate them and read them out as desired. Most outputs, however, update status variables of the same unit and/or become inputs to other units.

It should be borne in mind that all the data which are specific to com-

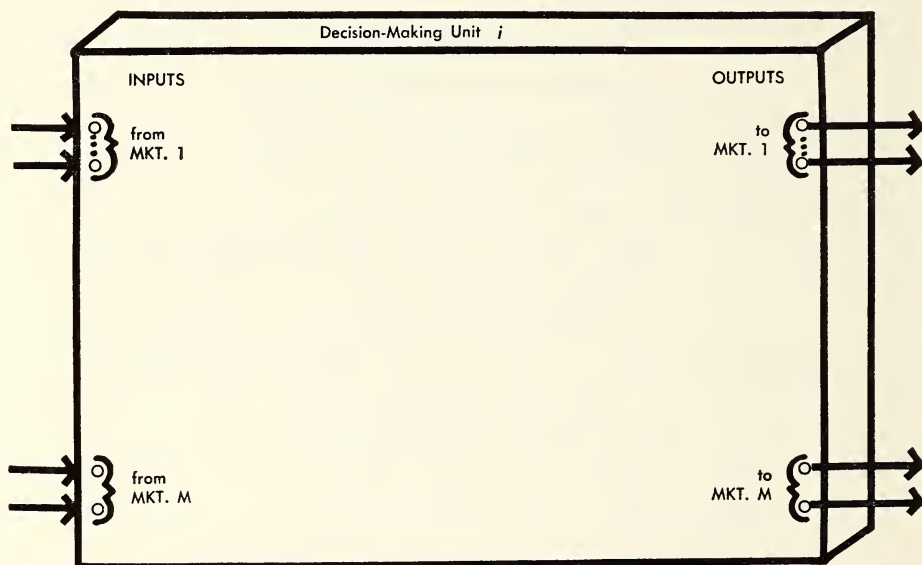


FIGURE 2. Flow diagram of a model of a decision-making unit.

ponents must, because of their volume, be stored on magnetic tape which is fed through the machine when the data are to be used. Operations are performed on each component or cluster of components in turn and the results written out on a new, updated tape. For example, if a woman with two children is designated on a given pass as having had a child, then she appears on the updated tape as a mother of three children and the information about the household or family to which she belongs is also altered correspondingly. If an output results in the breakup of a unit (e.g., divorce), two units appear on the updated tape where one appeared on the old tape. Moreover, there is no problem in using the output of one unit as an input to a unit which immediately follows it on the tape.

Using an output of one unit as an input to another unit which is randomly located on the tape is somewhat more difficult, since only data on a very limited number of units can be held in high-speed storage at any

one time. To illustrate one way of handling this, consider the simulation of marriage between two individuals (Chapter 6),¹ where we employ a two-stage process, in which units designated as *marrying* are then held in a *buffer* until they can be matched with spouses located elsewhere on the tape who have appropriate characteristics.

In many cases output-input linkage will involve both aggregation of outputs into a small enough set of numbers to be retained in high-speed storage and subsequent distribution of these aggregates to specific units. To give a specific illustration of this, let us suppose that the probability that a woman will marry during a given month is a function of various things such as her age, race, marital status, and also of the number and age-distribution of unmarried males. In this case the number of unmarried males in each age category would be counted during each pass of the household tape through the computer. The obtained distribution would be retained in high-speed storage and would be used until the next distribution became available.

It will have occurred to the reader that one of the most important types of economic behavior—purchase or sale—necessarily involves interaction between units. Here purchases could be regarded as outputs and the corresponding sales as inputs into the units doing the selling. One could conceive of using a two-stage process similar to the one we have used for marriage in which units designated as *purchasers* (or purchasers at a particular price) would be held in a buffer until they could be matched with *sellers* passing through on the tape. With a large number of products (or types of products) being purchased, however, this approach would necessitate far greater storage capacity than is available on present day electronic computers.

A more feasible alternative would be to aggregate purchases by industries (or some other grouping) and then use a Leontief type of input-output matrix to convert purchases into an appropriate set of sales by industries. The aggregate sales of any particular industry might then be allocated among firms according to characteristics of the individual firms.

This procedure is a possible one, but it involves certain practical difficulties well known to those who work with Leontief systems. If n industries or other subsectors are used, then n^2 coefficients are needed even under drastic assumptions. Not only is this a large number of coefficients if n is even moderately large, such as 100, but the information required to estimate the coefficients is difficult to obtain, since flows identified by source and origin are needed between every combination of subsectors. Moreover, 100 percent coverage of buyers and sellers seems to be required to estimate all of the coefficients. Since we intend operating with decision units, and therefore only a sampling of them, the usual input-output approach might be unnecessarily difficult. This is particu-

¹Orcutt *et al.*, *Microanalysis of Socioeconomic Systems: A Simulation Study* (New York: Harper & Bros., 1961).

larly the case since we wish to have a setup in which the linkage can be modified at frequent intervals as the structure of the economy undergoes change.

An alternative approach involving far fewer coefficients, less stringent data requirements, and permitting a richer and more realistic portrayal of market phenomena is the more traditional one of explicitly introducing markets. As it will be recalled, our second major class of components are markets and the specific function of these components is to transmit the outputs of decision-making units and to distribute them as inputs into other decision-making units.

Individual decision-making units that wish to sell something make known the prices and terms under which they are prepared to sell. These outputs of decision-making units are fed in as inputs into the appropriate markets. The operating characteristics of a given market summarize the offers to sell which it receives. The summarized information about price, delivery dates, etc., is then used in other operating characteristics of the market, and the outputs following from these operating characteristics are then fed to each potential buyer. The information fed to a potential buyer can be made to depend upon the potential buyer's location relative to that of the seller. If wished, potential buyers may receive only partial information, with some receiving more than others and possibly with some receiving incorrect price information.

Potential buyers receive the price and other information passed on to them through each market as inputs. These inputs are used in their operating characteristics and in some cases orders to buy are the resulting outputs. These orders enter as inputs into the appropriate markets. Operating characteristics of each market summarize and classify the orders by region of origin, price accepted, and so forth. Other operating characteristics of the market then use this summarized information about orders, along with information about each potential seller considered in turn, to distribute the orders among the potential sellers. Decision-making units respond to the orders which they receive as inputs and generate various outputs, among which in due course usually will be deliveries. These deliveries show up as inputs into an appropriate market, are transmitted and distributed by the market, and appear as outputs of the market and as inputs into the appropriate decision units. These decision units then make payments or promises to pay which are again distributed through the market to the appropriate sellers, and so forth.

A very sketchy flow diagram of a market is shown in Figure 3. The reader is to imagine connecting lines or cables from decision units to markets which feed the input variables into some of the operating characteristics of the markets. These market operating characteristics summarize and classify the data received and determine values of the summary and distributional variables. These status variables are then used in other market operating characteristics along with information from po-

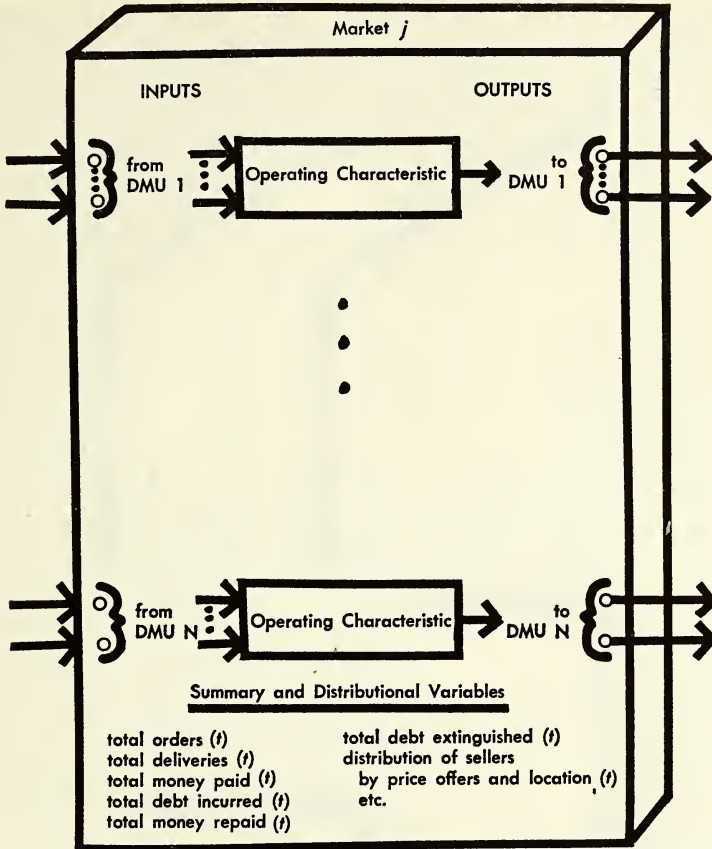


FIGURE 3. Incomplete flow diagram of a model of a market.

tential buyers or sellers. From these operating characteristics lines go to the various arrows indicating outputs from the market.

Figure 4 shows graphically how the outputs of the entire set of N decision-making units are transmitted and distributed by the M markets. The outputs of the decision-making units become inputs into the markets. After being summarized and distributed they appear as outputs of the markets. These market outputs then become inputs into the decision-making units again.

An essential characteristic of this type of model is that it proceeds in short discrete steps or periods. It is a recursive model; that is, the outputs of a unit in any period depend on prior inputs to the unit, so that there is no simultaneous interaction between units, and hence there are no simultaneous equations involving more than one unit at a time to be solved. This does not mean that units are conceived of as acting independently of each other, since the prior outputs of other units may be inputs to the unit in question; but it does mean that all interaction among units in the model is sequential rather than simultaneous. However, in-

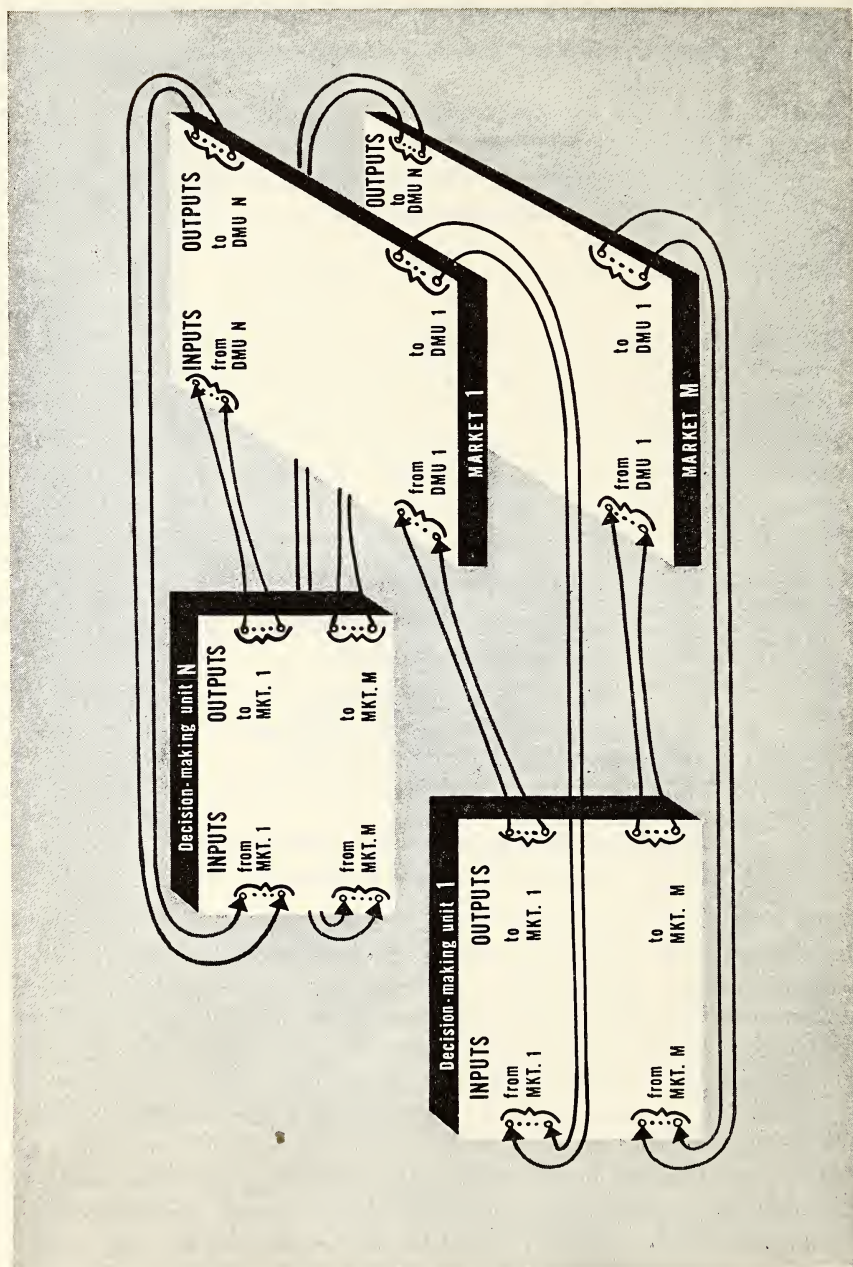


FIGURE 4. Incomplete flow diagram of a model of an economic system.

dividual components may contain simultaneous equations if this is desired.

If a recursive model is to be a good approximation to reality, the time period chosen must be short. Many responses, such as the response of purchase probabilities to a change in price or the reaction of marriage probabilities to a change in income, may be sequential in fact, but they will appear to be simultaneous if annual or other long-period data are analyzed. We have used the month as our basic period. An even shorter period, such as the week, might sometimes be preferable if the necessary data were available on this basis.

The implications of the model are worked out by simulation on a large electronic computing machine. Specifically, the demographic model of the household sector has been simulated on an IBM 704, but the basic procedure could be adapted to a Univac II or to one of the many more powerful successors to these computing giants. As a first step, a population of several thousand decision-making units is specified, the units being assigned initial values of their status variables in the proportions in which the corresponding characteristics appear in the units of the real socioeconomic system. These units with their status variables are listed in some arbitrary order on a magnetic tape which is fed into the machine. The simulation proceeds one month at a time. In each month each individual unit is considered in turn. A probability of occurrence is specified for each possible output of the unit by the relevant operating characteristics that are used in connection with the appropriate status variables and inputs to the unit at the beginning of the month. Whether the output occurs or not is determined by a random drawing from this probability distribution.

For example, suppose the simulation is started with the month of January 1960, and the first unit considered is an individual with the following characteristics (among others): male, white, age 34. We have already estimated the probability that an individual with these characteristics will die in this month at, say, 0.0002; i.e., this is the probability of death specified by the relevant operating characteristic. Now we make a random drawing from this distribution using random numbers generated by the machine for the purpose. There are two chances in 10,000 that the outcome of the draw will indicate *death* for this unit and 9998 that it will not. Depending on the outcome of the draw, the man either dies and is eliminated from the population (and from the household or other larger unit of which he may be a member) or he lives through the month.

In considering other aspects of a unit, there may, of course, be more than two possible outcomes. If the unit were a household, for example, the output *amount spent on durables* might have several possible values—\$0, \$1–100, \$101–200, and so forth.

When each possible output for each unit has been considered in this way, the first *pass*, or month, is complete. We enter the second month

with a population of units that is slightly different in both size and composition from the initial one. Some individuals have died or married, some babies have been born, some firms or households have been created or destroyed, many have altered their asset and debt positions, and so forth. The whole procedure is then repeated for the second month and for as many more as desired. To find out what has happened to the system at the end of, say, twelve months, we take a census and actually count the number of units with various characteristics or with combinations of characteristics.

In a model involving household units, firms, and financial institutions there would also be many markets. The operating characteristics of these markets and status variables relating to them would be stored in high-speed storage. As decision units place orders, for example, these orders would be summarized by operating characteristics of the appropriate market and the summarized information retained in high-speed storage. Then these orders will be distributed to the appropriate sellers as these decision-making units pass through the computer.

Simulation is certainly not the only conceivable approach to solving models of this type. Once an initial population and a complete set of operating characteristics were specified, one could try to deduce mathematically the probability distributions of the various aggregates at the end of a given number of periods. This certainly seems to be a straightforward approach, and it might yield a more precise knowledge of the solution; but it presents great practical difficulties. In order to compute the probability distributions associated with each aggregate of interest, it would be necessary to calculate the probability of each possible way of reaching each possible value of the aggregate and then sum these probabilities. However, the number of paths that might be followed by any single unit increases rapidly as the number of variables and time periods gets larger. Since every combination of each possible variation of path of each and every unit will correspond to a different path by which the system may generate aggregates, keeping track of all possible paths and their respective probabilities would seem to be an appalling computational task. The computational problem might be simplified by working with the first few moments of the distributions or by some other mathematical technique, and if someone tackles the model in this way we shall be grateful. In the meantime, simulation appears to be a more feasible alternative.

Even if a deductive solution were achieved, moreover, simulation would retain certain important advantages. It is possible to organize the simulation on a computer in such a way as to make it relatively easy to alter specific operating characteristics in the light of new knowledge or to make a new choice of aggregates to be observed. This means that it is possible to experiment with the model in order to find out how sensitive the results are to changes in the parameters. One can deliberately alter the values of given parameters on different runs, observe the

changes in a given aggregate, and then fit a multiple regression equation to get an estimate of the functional relationship between the aggregate and the values of these parameters.

Another illustration of the flexibility of the simulation approach, discussed below, is the possibility of substituting certain assumed values of the aggregates for the expected values used in the correction for small sample variation. This means that the model can be used to predict what will happen to some aggregates if assumed values are "plugged in" for others. The simulation approach is less likely to require restrictive assumptions in order to facilitate solutions. Lastly, but perhaps not least significantly, it is intelligible to persons of only modest mathematical sophistication.

The results obtained in a census of the units in the model at the end of a single simulation run may be expected to differ from what actually happens in the real system being simulated. The difference might arise from several sources. (1) The units of the real system may not behave according to the assumed operating characteristics. (2) Stochastic variation in the real system is not likely to be exactly duplicated in any given simulation. Even if events do behave like independent random drawings from stable probability distributions, the total number of events occurring in a finite population will differ on different trials. (3) There will be even greater stochastic variation in the model than in the real system, since the model consists of a much smaller number of units than the real system.

Let us assume for the moment that the first two sources of error can be ignored, that is, that we do have exact estimates of stable operating characteristics and that the number of units in the real system is so large that sampling variation in aggregate outputs is negligible. In simulating the system on a computer, however, it is still necessary, despite the immense power of modern computational equipment, to approximate the real system of millions of units with a reduced system containing thousands or at most tens of thousands of units. This should not affect the expected value of aggregates, but it substantially increases their sampling variability.

One solution would be to do quite a number of runs with the same initial population and the same operating characteristics and get a distribution of final results, from which could be estimated the expected numbers of units of given characteristics and the variances of these numbers. However, this would multiply the already substantial computing time required for a simulation and add greatly to the expense of the operation.

Another possibility is to operate our model containing a moderate number of units in such a way as to approximate a similar system containing, in effect, an infinite number of units. One way to do this would be to abandon our random drawing procedure and simply compute the expected outputs in each time period. Given the operating characteristics and the initial population cross-classified by all the relevant variables, we

could compute the number of outputs of each type expected to occur in each cell. Suppose, for example, that birth probabilities are taken to depend on age, sex, parity, and marital status, and that the expected number of births to married women aged 30-34 with four children is N ; then N women, randomly selected with respect to all other characteristics, would be subtracted out of the cell "married, female, age 30-34, parity 4" and moved in the next period to "married, female, aged 30-34, parity 5." This is essentially the method used by the Bureau of the Census to project population. The method is quite feasible when only two or three inputs and outputs are involved, but it becomes very cumbersome as the number of inputs and outputs increases. It is necessary, at each point in the computation, to collect all the units which are the same with respect to certain characteristics, compute the expected number to be moved out of this group, and then pick the units to be moved at random with respect to all other characteristics. Doing this on a computer would involve repeated searching of the tape, which is a relatively time-consuming operation, and it would be much more expensive and less flexible than the sequential treatment we accord to each component.

Without abandoning our basic approach, we have introduced a device which at least gets rid of small sample variation in the aggregate outputs of the model. Every time a random drawing is made, we have the machine record the expected outcome as well as the actual outcome. These expected outcomes are accumulated, so that at the end of a pass we know the expected number (or value) of births, deaths, expenditures, and other aggregates as well as the numbers generated by the random drawing procedure. The machine keeps track of the cumulative discrepancies between the expected and the generated aggregates from one period to the next. The cumulated discrepancies in each period are then used to adjust the probabilities used in the next period's simulation in such a way as to keep the cumulated discrepancies near zero. In the calculation of expected numbers only unadjusted probabilities are used.

This device nearly eliminates the problem of sampling variability in the aggregates. It does not change the fact that, unless this device is used for each cell of a joint distribution, there will be sampling variation in the number of units which fall in the individual cells. For example, if one wanted to obtain a distribution of households by region, size, age of head, income, and various assets, it would still be necessary to do several runs in order to get estimates of the mean and variance of each cell (even if one were to ignore the errors of estimate in the operating characteristics). Moreover, there are instances in which the units of the real system are not very numerous and where one might conceive of there being sampling variation in the real system as well as in the model. In this instance one might want to do several simulation runs with the same units and the same operating characteristics in order to obtain estimates of this sampling variability.

Estimating the operating characteristics associated with all the relevant outputs of all the different kinds of units which make up our economic system is a very, very large order. Fortunately, the problem can be broken down into major parts, and separate research efforts can be mounted on the similar components within each sector. Since the flow-of-funds system of national accounts was consciously organized around transactors (i.e., economic decision-making units), the sector classifications used in the flow-of-funds system are convenient for our purposes. Use of this sector classification establishes a link between the aggregates predicted by the model and a substantial body of readily available data, thus increasing the potential importance of the predictions made by the model and providing historical series that can be used in checking the model. The flow-of-funds sectors are:

Consumers	State and local governments
Corporate business	Banking system
Nonfarm noncorporate business	Insurance
Farm business	Other institutional investors
Federal government	Rest of the world

A general model would include all these sectors—possibly with regional or other subsectors. Models of single sectors or groups of sectors are useful in themselves, however, in generating conditional predictions about the behavior of units in some sectors of the system on specific assumptions about the aggregate behavior of units in other sectors.

So far we have concentrated our attention on the consumer sector. This seemed like a good place to start, not only because of the interests of the authors, but because the consumer sector consists of a very large number of units about whose behavior we are beginning to amass a substantial body of data which has not yet been satisfactorily integrated into a model of the economic system. Since the consumer sector seems to abound with nonlinearities, it is difficult to handle in models in which linearity assumptions are crucial. Thus there is here a particular need for a model which does not stipulate restrictions on the functional forms involved.

The basic unit of the consumer sector is, of course, the individual, but many consumption decisions are apparently made jointly by groups of individuals; for example, married couples, families, households, and spending units. The following is a partial list of possible outputs of units in the consumer sector:

Birth	Student status	Grants and donations
Death	Purchases of	Insurance premiums
Marriage	Durable goods	Change in
Divorce	Nondurable goods	Mortgage debt
Labor-force status	Services	Nonmortgage debt
Occupation	Houses	Liquid assets
Employment status	Taxes paid	Nonliquid assets

A list of possible status variables and inputs to consumer units might get very long. The following will give some idea of the type of detail about each unit that might practicably be carried along as status variables or might be generated or fed in and utilized as needed:

Specific to Unit

Household identification	Mortgage debt
Number, race, and sex of adults	Nonmortgage debt
Number of male children	Liquid assets
Number of female children	Nonliquid financial assets
Region	Housing stock
Population density (city size)	Durables stock
Education of male adult	Income of male adult
Occupation of male adult	Interval since marriage, divorce, or death of spouse
Education of female adult	Age of male adult
Occupation and labor force status of female adult	Age of female adult
Household income	Ages of all children
Moving average of household income	

Nonspecific to Unit

Time
Tax laws
Credit availability
Prices
Employment opportunities

A schematic diagram of a model of a "nuclear" family is shown in Figure 5. For our purposes a nuclear family might consist of a married couple and their own children under 19 years of age. In this case we have individual components embedded within a larger component, the nuclear family. Some status variables and operating characteristics are thought of as being specifically connected with individuals. The output generated by these operating characteristics might include death, birth, entrance into the labor force, acceptance of a marriage offer. Other status variables and operating characteristics are associated with the family as a whole. The outputs generated by these might include expenditures, management of assets, incurring of debts. The various operating characteristics update the status variables and generate the outputs of the family and individual components.

Microanalytic models are ideally set up to use data and relationships that apply at a microlevel. This is one of the major advantages of this type of model and should be fully exploited. Nevertheless, certain things should be kept in mind. In some cases data needed to determine appropriate microoperating characteristics may not be available. In other cases the data may be available, but the particular need may not justify the added complexity or computing effort required fully to utilize the microinformation. However, there is nothing about microanalytic models or their simulation that restricts the model builder to the use of microdata or microrelations. He has the added opportunity of building at the

microlevel, but he retains an equal facility of incorporating the use of aggregative data and relations.

Our strategy is to build and test as completely as possible at the microlevel, but then to use aggregative data and relationships to complete our models and to bring them into alignment with historical aggregative data. This puts as much of the burden of testing and formulation as is feasible at the microlevel, where it belongs. It retains the use of aggregative data as fully as possible for final testing and alignment of the overall model. As data availability improves, and as our knowledge grows, microanalytic-type models of the economic system may, and in fact should, place less and less reliance on aggregative data and relationships.

However, this should be a gradual evolution, and, since the final reward of these models will usually be in predicting things about aggregates, it follows that final testing against aggregative data will always be necessary. Hopefully, such testing will not require extensive gross adjustments aimed at bringing the model into line with the aggregative data used in testing.

A detailed discussion of our attempts to estimate the operating characteristics for some of the units in the consumer sector and to describe the simulation of a limited model of household demographic behavior is presented in Chapters 3 to 19 of the book from which this article was extracted.² It is our hope that these efforts will encourage other social scientists to work along the same general lines and to help fill in the missing pieces—both in the consumer sector and in other sectors.

Microanalytic models of the type suggested here can increase the range of our predictions in two ways. First, through the incorporation of wider ranges of behavior within the models, predictions could be made in areas with which existing models of our socioeconomic system do not deal. Secondly, these predictions would relate to both single and multivariate distributions, all quickly accessible in tabular or graphical form by spot interrogation of the updated tapes in much the same way that the human population is interrogated during a census taking. Such models would also facilitate and improve prediction about socioeconomic aggregates by providing a method of bringing to bear knowledge about the elemental decision-making units that make up a socioeconomic system.

Such models could be used either for short-run or long-run forecasting by appropriate selection of initial conditions and by altering the number of periods the model is run. They could be used either for unconditional forecasting or for predictions of what would happen given specified external conditions and governmental actions.

Models of the type suggested here could facilitate and improve test-

² Orcutt *et al.*, *op. cit.*

ing of hypotheses about elemental units by permitting testing of hypotheses at any level of aggregation. Such models also would improve the testing of such hypotheses by keeping the interrelated nature of the system in the consciousness of the investigator, and by helping him to take it into account satisfactorily.

The role of microanalytic models in guiding selection of research efforts would be similar in nature to that provided by other models of the socioeconomic system. They permit the researcher to see how small pieces can be fitted together and to see where there are serious gaps or weaknesses. They enable him to produce a small piece that will contribute effectively to a useful whole. Since most research can be done effectively only in fairly small pieces, this is important. The main advantage of this sort of model in providing guidance in selection of research effort lies in the fact that the basic components are elemental decision-making units and other microcomponents of a sort not yet effectively incorporated into other available models of our socioeconomic system.

One of the major objectives of the approach taken is to provide an instrument for consolidating past, present, and future research efforts of many individuals in varied areas of economics and sociology into one effective and meaningful model; an instrument for combining survey and theoretical results obtained on the microlevel into an all-embracing system useful for prediction, control experimentation, and analysis on the aggregate level. The possibilities of such a system tempt the imagination. In the ultimate analysis, the predictive value of such a system will depend entirely on the quality of the research effort that enters into formulation of its components.

A Model of Retail Ordering and Pricing by a Department Store*

.....

R. M. CYERT, J. G. MARCH, and C. G. MOORE†

THE METHODS AND PRACTICES BY WHICH RETAIL FIRMS MAKE ORDERING and pricing decisions are discussed in most retailing books and in many introductory marketing texts. For the most part, however, the descriptions outlined are of a general nature. The focus has been on the broad aspects of the problems rather than the specifics of any given ordering pricing process. This paper describes a simulation model of ordering and pricing behavior in a specific department of a large retail department store. The predictions generated by the model are shown to bear a close relation to the actual observed behavior.

The organization chosen for intensive study is one department in a large retail department store. The firm involved is part of an oligopolistic market consisting (for most purposes) of three large downtown stores. Each of the firms involved also operates one or more suburban stores; but the focus of this study, and the bulk of the sales for each store, is the downtown market. The firm is organized into several merchandising groups. Each of these groups has several departments. The firm, in total, has more than 100 major departments. We have studied, with varying degrees of intensity, the price and ordering decisions in about a dozen of the firm's departments. From these dozen we have chosen one for intensive investigation, and the specific model reported here is literally a model of decision making in that specific department. In our judgment, the decision processes we report for the one department could be generalized with trivial changes to other departments in the same merchandising group and could be generalized with relatively modest changes to most other departments outside the immediate group.

* Copyright 1962 by Prentice-Hall, Inc. This paper is based on a chapter in R. M. Cyert and J. G. March, *A Behavioral Theory of the Firm*, to be published by Prentice-Hall, Inc., in 1963.

† All of the Carnegie Institute of Technology.

Because of the great similarity in operation among department stores, the model probably represents most aspects of decision making in a retail department store in general.

We present the model at two levels of specificity. In the first section we outline the decision process in the organization in rather general terms. In the second section we elaborate some of the specific decision rules in order to provide specific, explicit predictions of decisions.

GENERAL VIEW OF PRICING AND ORDERING IN A RETAIL DEPARTMENT STORE

The organization studied makes relatively independent pricing and ordering decisions. There are loose connections between them, but for the most part they are made with reference to different goals and different stimuli. Although we will want to elaborate the goals of the organization somewhat when we turn to specific decision rules, we can describe two general goals that the department pursues. (1) The department expects (and is expected by the firm) to achieve an annual sales objective. (2) The department attempts to realize a specified average mark-up on the goods sold. Organizational decision making occurs in response to problems (or perceived potential problems) with respect to one or the other of these goals.

Sales Goal

The general flow chart for decision making with respect to the sales goal is indicated in Figure 1. The organization forms sales "estimates" that are consistent with its sales goal and develops a routine ordering plan for advance orders.¹ These orders are designed to avoid overcommitment, pending feedback on sales. As feedback on sales is provided, results are checked against the sales objective. If the objective is being achieved, reorders are made according to standard rules. This is the usual route of decisions. We will elaborate it further below.

Suppose, however, that the sales goal is not being achieved. Under such circumstances a series of steps are taken. First, the department attempts to change its environment by negotiating revised agreements with either its suppliers or other parts of its own firm or both. Within the firm, it seeks a change in the promotional budget that will provide greater promotional resources for the goods sold by the department. Outside the firm, the department seeks price concessions from manufacturers that will permit a reduction in retail price. If either of these attempts to relax external constraints is successful, reorders are made according to appropriately revised rules.

¹ This statement may not portray the process accurately. During the period we observed it was not possible to determine the interactions between the sales estimates and the goals. They always tended to be consistent with each other but it was difficult to determine the extent to which an implicit goal of "equal or exceed last year's sales" influenced the estimates.

Second, the department considers a routine markdown to stimulate sales generally and to make room for new items in the inventory. As we will indicate below, the department ordinarily has a pool of stock available for markdowns and expects to have to reduce mark-up in this way on some of the goods sold. It will attempt to stimulate all sales by taking some of these anticipated markdowns. Once again, if the tactic is successful in stimulating sales, reorders are made according to slightly revised rules.

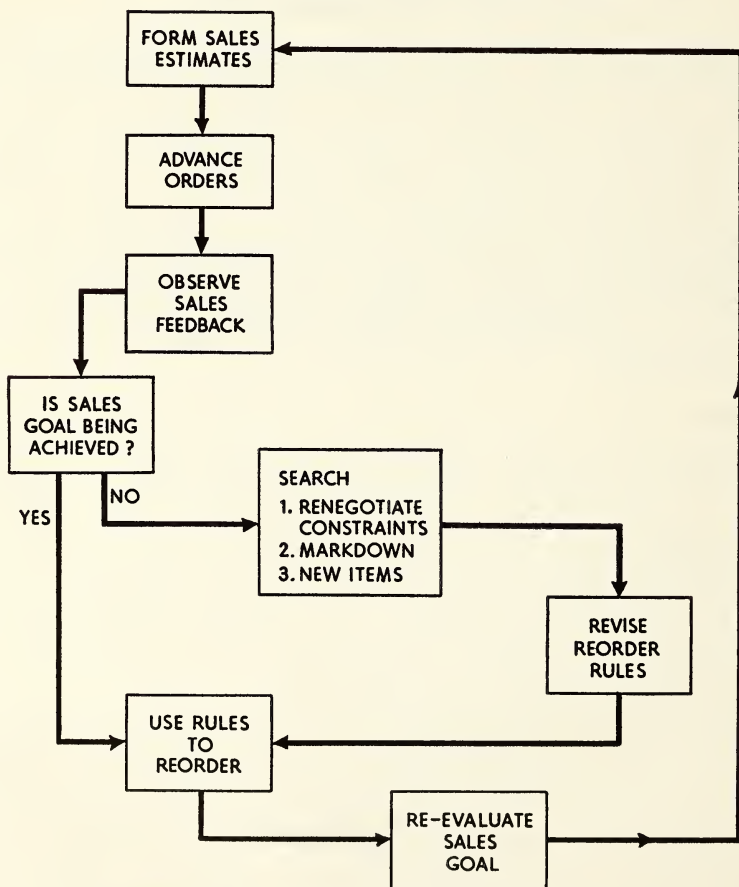


FIGURE 1. General form of reaction to sales goal indicators.

Third, the department searches for new items that can be sold at relatively low price (but with standard mark-up). Most commonly, such items are found when domestic suppliers are eliminating lines or are in financial trouble. A second major source is in foreign markets.

In general, the department continues to search for solutions to its sales problems until it finds them. If the search procedures are successful, all

goes well. In the long run, however, it may find a solution in another way. The feedback on sales not only triggers action, it also leads to the reevaluation of the sales goal. In the face of persistent failure to achieve the sales goal, the goal adjusts downward. With persistent success it adjusts upward.

Mark-up Goal

The flow chart in Figure 2 outlines the departmental reaction with respect to the mark-up goal. It is analogous to, but somewhat different in

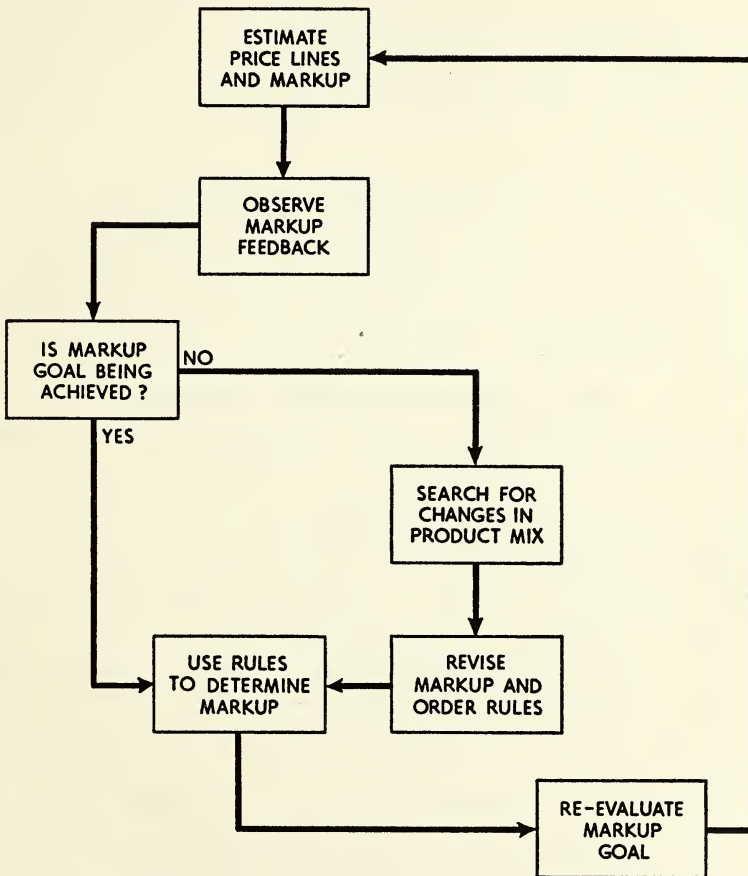


FIGURE 2. General form of reaction to markup goal indicators.

impact from, the reaction to the sales goal. On the basis of the mark-up goal (and standard industry practice), price lines and planned mark-up levels are established. Feedback on realized mark-up is received. If it is consistent with the goal, no action is taken and standard decision rules are maintained.

If the mark-up goal is not being achieved, the department searches for

ways in which it can raise mark-up. Basically the search focuses on procedures for altering the product mix of the department in the direction of increasing the proportion of high mark-up items sold. For example, the department searches for items that are exclusive, for items that can be obtained from regular suppliers below standard cost, and for items from abroad. Where some of the same search efforts led to price reduction (and maintenance of mark-up), when stimulated by failure on the sales goal, here they lead to maintenance of price and increase in mark-up. At the same time, the organization directs its major promotional efforts toward items on which high mark-ups can be realized. In some instances, the department has a reservoir of solutions to mark-up problems (e.g., pressure selling of high mark-up items). Such solutions are generally reserved for crises and are not viewed as appropriate long run solutions. Finally, as in the case of the sales goal, the mark-up goal adjusts to experience gradually.

We think the general process reflected in Figures 1 and 2 correctly represent the decision process in the firm under examination. They do not, however, yield specific, precise predictions. Detailed models for most (but not all) of major decisions need to be developed and compared with output from the organization studied. This is the task undertaken in the next section of this paper.

DETAILS OF THE ORDERING PROCESS

As has already been suggested, the ordering decision is essentially based on feedback from sales experience. Virtually no explicit calculation of the probable behavior of competitors is made; and although expectations with respect to sales are formed, every effort is made to avoid depending on any kind of long run forecast. Ordering decisions are designed to satisfy two major goals: (1) limit markdowns to an acceptable level; and (2) maintain inventory at a reasonable level.

The firm divides ordering decisions into two classes—advance (initial) orders and reorders. Each is dependent on a different set of variables and performs a different function. Advance orders allow the firm (and its suppliers) to avoid uncertainty by providing a contractual environment. They also account for the bulk of the total orders. Reorders account for only a minority of the total orders; but they provide virtually all the variance in total orders. Insofar as the ordering decision is viewed primarily with respect to the total quantity of orders, reorders are much more important than advance orders in fixing the overall level.

Advance Orders

Advance orders represent the base level of orders for the department. The size of the advance order depends on two things. One of these is the estimated sales. The other is a simple estimate of the variance in sales.

In a general way, the apparent motivation with respect to advance orders is to set the commitment at such a level that the base output alone will be greater than sales only if extreme estimation errors have been made. Thus, refinements in estimation are not attempted and simple estimating procedures are used, modified somewhat by special organizational needs only remotely related to the issue of accuracy.

The Estimation of Sales. The store operates on a six-month planning period. The individual department estimates dollar sales expected during the next six months. At the same time, sales are estimated on a monthly basis for each product class over the six-month period. Since the accuracy of the estimate is not particularly critical (at least within rather broad limits) for the overall level of orders, we consider the organizational setting in which the estimate is made in order to understand the decision rules. A low forecast, within limits, carries no penalties. The forecast cannot, however, be so low relative to past history that it draws criticism (as being unrealistic) from top management. The major limit on the low side is simply that the forecast be within a reasonable distance of past achievements. Limits on the high side are specified by two penalties for making a forecast that is not achieved. First, achievement of forecasts is one of the secondary criteria for judging the performance of the department. Although the department cannot significantly affect the sales goal by underestimation, it can to a limited extent soften criticism (for failure) by anticipating it. Second, an overestimate will result in over-allocation of funds to the buyer. If he is unable to use the funds, he is subject to criticism. As a result, the buyer's estimate tends to be biased downward.

The primary data used in estimating sales are the dollar sales (at retail prices) for the corresponding period in the immediately previous year. Although these data are commonly adjusted slightly for "unique" events, the adjustments are not highly significant. The following naive rule predicts the estimates with substantial accuracy:²

RULE 1: The estimate for the next six months is equal to the total of the corresponding six months of the previous year minus one-half of the sales achieved during the last month of the previous six-month period.

From the point of view of ordering decisions, the more critical estimates are those for the individual months. The monthly figures are used directly in determining advance orders for the individual seasons. The estimation procedure for a particular product class is as follows:

RULE 2: *For the months of February, March, and April.* Use the weekly sales of the seven weeks before Easter of the previous year as the estimate of the seven weeks before Easter of this year. In the same way, extend the sales

² We do not mean to imply that the department consciously uses such a rule. Although the rule was inferred from a study of actual behavior, the head of the department did not describe his estimation rule in these terms.

of last year for the weeks before and after the season to the corresponding weeks of this year.

RULE 3: *For the months of August and September.* The same basic procedure as in Rule 2 is followed with the date of the public schools' opening replacing the position of Easter. The opening dates of county and parochial schools also are significant. If these dates are far enough apart, the peak will be reduced, but the estimate for the two months will still represent the total sales of the corresponding two months of the previous year.

RULE 4: *For the months of May, June, October, November, and December.* Estimated sales for this year equal last year's actual.

RULE 5: *For the months of January and July.* Estimated sales for this year equal one-half of last year's actual rounded to the nearest \$100.

This set of simple rules provides an estimate of sales that is tightly linked to the experience of the immediately previous year with a slight downward adjustment.

The Seasonal Advance Order Fraction. The department distinguishes four seasons—Easter, Summer, Fall and Holiday. The seasons, in fact, do not account exhaustively for all months, but they account for most of the total sales. Estimates of sales are established on the basis of the monthly estimates. These estimates do not necessarily include all months in the season. The following estimation rules are used:

Easter: Cumulate sales for the seven weeks before Easter

Summer: Cumulate sales for April, May, June, and $\frac{1}{2}$ of July

Fall: Cumulate sales for $\frac{1}{2}$ of July, August, and September

Holiday: Cumulate sales for October, November and December

These cumulations give a seasonal sales estimate for use in establishing advance orders. Once such an estimate is made, some fraction of the estimated sales are ordered.

Advance orders generally offer some concrete advantages to the department. Greater selection is possible (some goods may not be available later) and some side payments may be offered by the producer (e.g., credit terms, extra services). The department exploits these advantages by ordering a substantial fraction of its anticipated sales in advance. But an attempt is made to limit the advance order fraction to an amount that would be sold under even an extreme downward shift in demand.

The advance order fraction is subject to learning on the part of the organization and reflects primarily differences among the seasons with respect to sales variability and uniqueness. The greater the susceptibility of seasonal sales to exogenous variables (e.g., weather), the lower the fraction. The more specialized the merchandise sold during a season (i.e., the greater the difficulty of carrying it over to another season), the lower the fraction. At the point in time at which we observed the organization, the fraction (estimated from interviews and analysis of data) and the timing of advance orders for the four seasons were as follows:

Seasons	% of Estimated Sales Placed in Advance Orders	Time Order Made
Easter.....	50	January 15-20
Summer.....	60	March 10-15
Back-to-school.....	75	May 20-25
Holiday.....	65	September 20-25

In general we expect this simple model to predict quite well, being off only when the buyer makes *ad hoc* adjustments. Our observations lead us to believe this will not happen frequently.

Reorders

For all practical purposes, reorders control the total orders of the department. As the word implies, a reorder is an order for merchandise made on the basis of feedback from inventory and sales. Because of lead-time problems, much of the feedback is based on early season sales information. Thus, the timing of a reorder depends on the length of time to the peak sales period as compared with the manufacturing leadtime required.

Reorder Rules. Reorders are based on a re-estimate of probable sales. Data on current sales are used in a simple way to adjust "normal" sales. The reorder program specifies reorders for a given type of product class as a result of a simple algebraic adjustment.

Let T = the total period of the season.

τ = the period of the season covered by the analysis.

S_{ir} = this year's sales of product class i over τ .

S'_{ir} = last year's sales of product class i over τ .

$S'_{i(T-\tau)}$ = last year's sales of product class i over $T - \tau$.

I_i = available stock of i at time of analysis including stock ordered.

M = minimum amount of stock of i desired at all times.

$O_{i(T-\tau)}$ = reorder estimate.

Then

$$O_{i(T-\tau)} = \left[\frac{S_{ir}}{S'_{ir}} \cdot S'_{i(T-\tau)} + M \right] - I_i$$

If $O_{i(T-\tau)} \leq 0$ no reorders will be made. In addition, orders already made may be cancelled, prices may be lowered, etc. to reduce the presumed overstocking. Such an analysis would be made for each product class. The figure that results is a tentative one subject to minor modifications in the light of anticipated special events.

Open-to-Buy Constraint

The firm constrains the enthusiasms of its departments by maintaining a number of records relevant to ordering decisions. One of the more conspicuous of such records is the "open-to-buy." The open-to-buy is, in

effect, the capital made available to each department for purchases. The open-to-buy for any month is calculated from the following equation:

$$B_{\tau} = (I_{\tau+1}^* - I_{\tau}) + S_{\tau}^*$$

where B_{τ} = open-to-buy for month τ

$I_{\tau+1}^*$ = expected inventory (based on seasonal plans) for beginning of month $(\tau + 1)$.

I_{τ} = actual inventory at beginning of month τ .

S_{τ}^* = expected sales.

The department starts each month with this calculated amount (B_{τ}) minus any advance orders that have already been charged against the month. Any surplus or deficit from the preceding month will increase or decrease the current account, as will cancellation of back orders or stock price changes.

While the open-to-buy is a constraint on ordering, the exact constraint is hard to specify. On the one hand, so long as the rules above are followed and the environment stays more or less stable, the open-to-buy constraint rarely will be violated. From this point of view, the open-to-buy is simply a long run control device enforcing the standard reorder procedure and alerting higher levels in the organization to significant deviations from such procedure. On the other hand, the constraint is a flexible one. It is possible for a department to have a negative open-to-buy (up to a limit of approximately average monthly sales). Negative values for the open-to-buy are tolerated when they can be justified in terms of special reasons for optimistic sales expectations (e.g., when feedback has been affected by a transit strike).

The open-to-buy, thus, is less a constraint than a signal—to both higher management and the department—indicating a possible need for some sort of remedial action.

DETAILS OF THE PRICING PROCESS

The firm recognizes three different pricing situations: (1) normal pricing, (2) sales pricing, and (3) markdown pricing. The first two situations occur at regularly planned times. The third situation is a contingent situation, produced by failure or anticipated failure with respect to organizational goals. In each pricing situation the basic procedure is the same, the application of a mark-up to a cost to determine an appropriate price (subject to some rounding to provide convenient prices).

The bulk of sales occur at prices set by either of the regular standard procedures (normal and sales pricing). Markdown pricing is one of the main strategies considered when search is stimulated. During the time period we observed, the demand was strong enough to permit fairly consistent achievement of the department's pricing goal—an average realized mark-up in the neighborhood of 40%. As a result, we did not observe actual situations in which the pressure to reduce prices stemming from

inventory feedback conflicted with the pressure to maintain or raise mark-up stemming from overall mark-up feedback.

Normal Pricing

Normal pricing is used when new output is accepted by the department for sale. As has already been observed, the problem of pricing is simplified considerably by the practice of price lining. In effect, the retail price is determined first and then output that can be priced (with the appropriate mark-up) at that price is obtained. Since manufacturers are also aware of the standard price lines, their products are also standardized at appropriate costs.

For each product group in the firm there is a normal mark-up. Like the seasonal advance order fraction, mark-up is probably subject to long run learning. It varies in a general way from product group to product group according to the apparent risks involved, the costs of promotion or handling, the extent of competition, and the price elasticity. But in any short run the normal mark-up is remarkably stable. The statement is frequently made in the industry that mark-ups have remained the same for the last 40 or 50 years.

Standard Items. In the particular department under study, normal mark-up is 40%. By industry practice, standard costs (wholesale prices) ordinarily end in \$.75. By firm policy, standard prices (retail prices) ordinarily end in \$.95. Thus, all but two of the price levels are in accord with the following rule:

"Divide each cost by 0.60 (i.e., $1 - \text{mark-up}$) and move the result to the nearest 95¢." The results of this rule and the effective mark-ups are shown in Table 1.

TABLE 1
STANDARD PRICES

<i>Standard Costs (Dollars)</i>	<i>Standard Price (Dollars)</i>	<i>Effective Markup (Percent)</i>
3.00.....	5.00	40.0
3.75.....	5.95	37.0
4.75.....	7.95	40.2
5.50.....	8.95	38.5
6.75.....	10.95	38.3
7.75.....	12.95	40.1
8.75.....	14.95	41.5
10.75.....	17.95	40.0
11.75.....	19.95	41.0
13.75.....	22.95	40.0
14.75.....	25.00	41.0
18.75.....	29.95	37.4

Exclusive Items. In some cases, the department obtains items that are not made available to competition. For such products and especially where quality is difficult to measure, the buyer prices higher than the standard. The pricing rule is as follows:

When merchandise is received on an exclusive basis, calculate the standard price from the cost, then use the next highest price on the standard schedule.

Import Items. Presumably because they are frequently exclusive items, because of somewhat greater risks associated with foreign suppliers, and because of the generally lower costs of items of foreign manufacture for equal quality, the department increases the mark-up for imported items. For the product class studied the standard accepted increase over normal is 50% (which gives a target mark-up of 60%). This leads to the following rule for pricing imports:

Divide the cost by .40 (i.e., 1 - mark-up) and move the result to the nearest standard price. If this necessitates a change of more than 50 cents, create a new price at the nearest appropriate ending (that is, \$.95 or \$.00).

Regular Sale Pricing

We can distinguish two situations in which normal pricing is not used. One is during the regular sales held by the firm a few times during the year. The other is when the department concludes a markdown is needed to stimulate purchases, or to reduce inventory levels. In this section we consider the first case. As in the case of normal pricing, sales pricing depends on a series of relatively simple rules. In almost all cases sales pricing is a direct function of either the normal price (i.e., there is a standard sales reduction in price) or the cost (i.e., there is a sales mark-up rule). Both the figures on reduction and the sales mark-up are conventional, subject perhaps to long run learning but invariant during our observations. The general pricing rules for sales operate within a series of constraints that serve to enforce minor changes either to ensure consistency within the pricing (e.g., maintain price differentials between items, maintain price consistency across departments), or to provide attractive prices (e.g., use prices a few cents below the dollar level or, when feasible, "alliterative" prices such as \$3.30, \$4.40, etc.).

General Constraints. The department sets its sales prices within the framework of five policy constraints. These constraints have not changed in recent years and are viewed by the organization as basic firm policy. They are only rarely subject to review.

1. If normal price falls at one of the following lines, the corresponding sale price will be used:

<i>Normal Price</i>	<i>Sale Price</i>
\$1.00	\$0.85
1.95	1.65
2.50	2.10
2.95	2.45
3.50	2.90
3.95	3.30
4.95	3.90
5.00	3.90

2. For all other merchandise, there must be a reduction of *at least* 15% on items retailing regularly for \$3.00 or less and *at least* 16 $\frac{2}{3}$ % on higher priced items.
3. All sales prices must end with 0 or 5.
4. No sale prices are allowed to fall on any of the price lines normal for the product group concerned.
5. Whenever there is a choice between an ending of 0.85 and 0.90, the latter ending will prevail.

Departmental Decision Rules. Subject to the general policy constraints, the department is allowed a relatively free hand. Since the policy constraints do not uniquely define sales pricing, it is necessary to determine the departmental decision rules. These rules are indicated in detail in the flow charts in Figures 3 and 4.

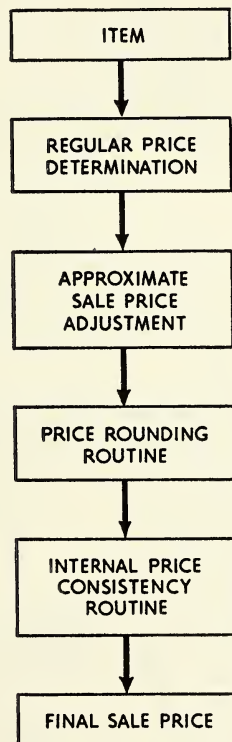


FIGURE 3. Major subroutine of sale pricing decision.

Markdown Pricing

In our earlier discussions of organizational decision making, we have suggested that a general model of pricing and ordering must distinguish between the ordinary procedures (that account for much of the bulk of the actual decisions made) and the special search procedures that are triggered by special circumstances. We have already seen how such procedures enter into the determination of orders for the organization

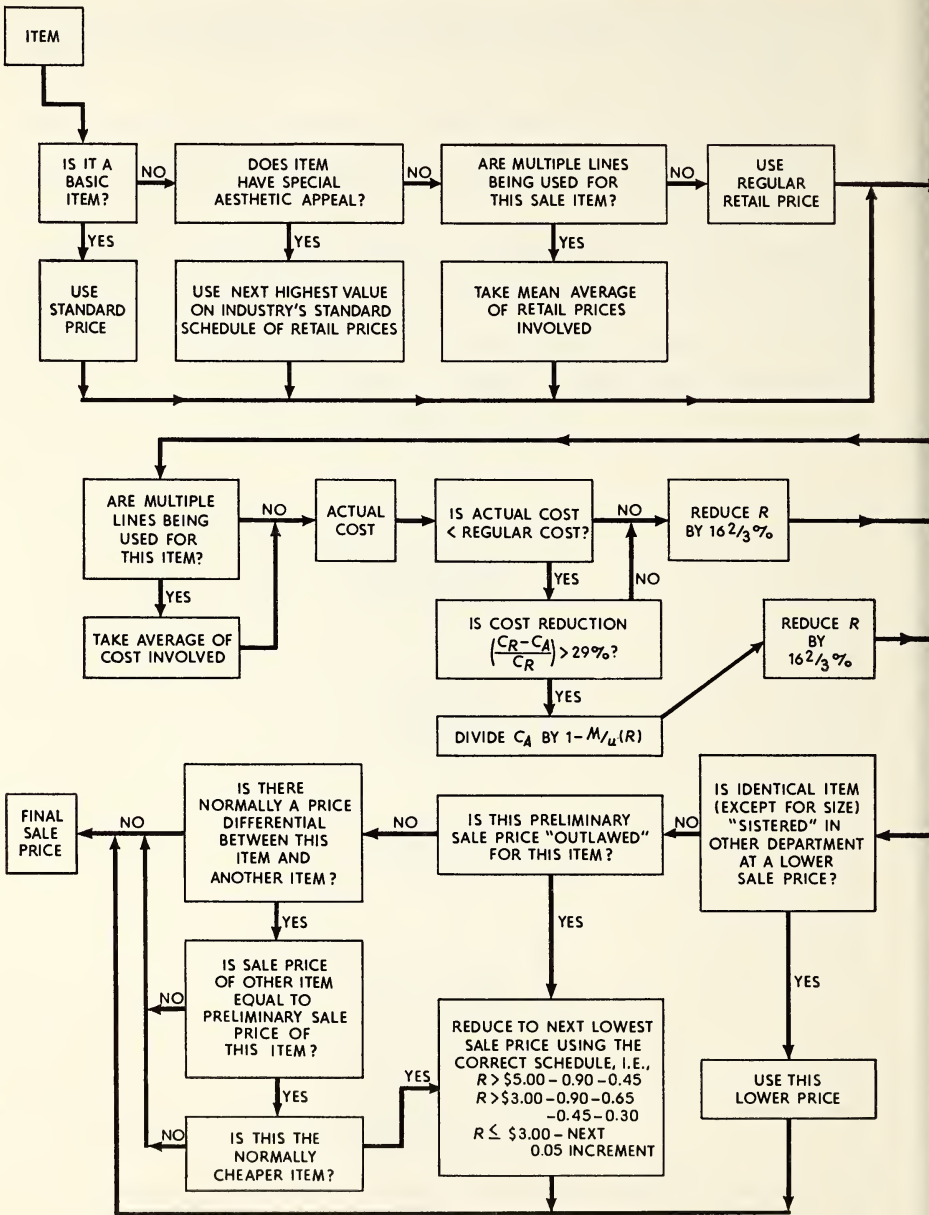
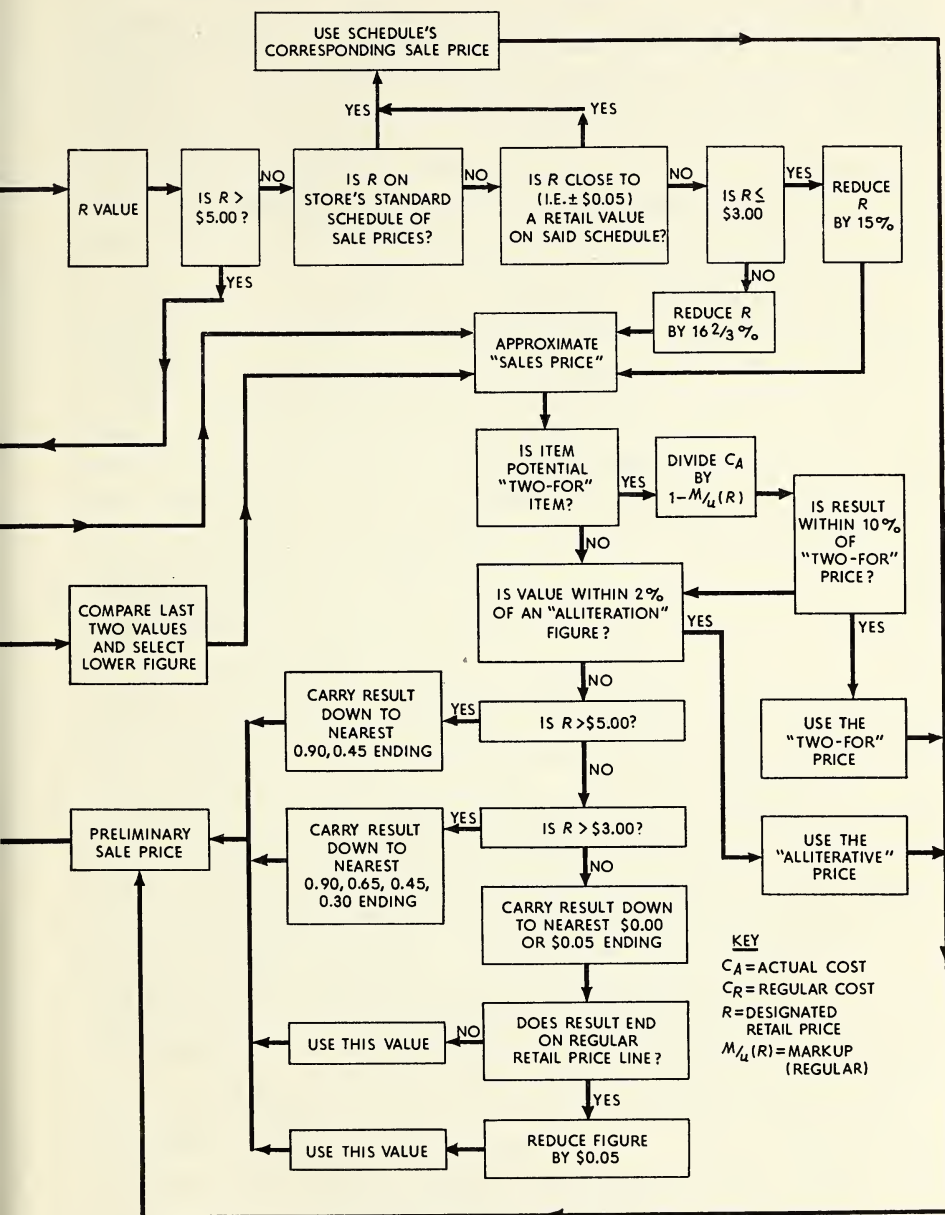


FIGURE 4. Flow chart for sale pricing decision.

studied. We turn now to search and "emergency" behavior on the price side. Price is the major adaptive device open to the department in its efforts to meet its mark-up goal, maintain sales, maintain inventory control, and in general meet the demands of other parts of the organization. We have already indicated how an increase in mark-up (e.g., on imports



and special items) is used by the department. We turn now to markdowns.

With respect to markdowns, the department has two decisions to make: When? And how much? In a general sense, the answer to the first question—the question of timing—is simple. The organization reduces price

when feedback indicates an unsatisfactory sales or inventory position. The indicators include the inventory records, sales records, or physical inventory reports on competitive prices, and the open-to-buy report. Of secondary importance are markdowns because of product properties (e.g., defective items). With respect to the amount of markdown, the organization has a set of standard rules. They have developed over time in such a way that their rationale can only be inferred in most cases. But a consistent characteristic of the rules is the extent to which they avoid pricing below cost except as a last resort.

Timing of Markdowns. Occasions for markdowns are primarily determined by feedback on sales performance. There are three general overstock situations that account for the majority of the markdowns. They are: (1) normal remnants, (2) overstocked merchandise, and (3) unaccepted merchandise. Normal remnants are the odd sizes, less-popular colors and less-favored styles remaining from the total assortment of an item which sold satisfactorily during the season. Overstocked merchandise includes items that have experienced a satisfactory sales rate but for which the buyer was overly optimistic in his orders. As a result, the season ends with a significant inventory that is well balanced and includes many acceptable items. Unaccepted merchandise represents merchandise that has had unsatisfactory sales. The sales personnel try to determine during the season whether the lack of acceptance is due to overpricing or poor style, color, etc. The distinction is usually made by determining whether or not the item has been ignored. If it has, the latter cause is usually inferred. If the item gets attention but low sales, the inference is that the price is wrong. In addition, there are a number of quantitatively less important reasons for considering markdowns. For example, the firm will meet competition on price (if a check indicates the competitor's price is not a mistake). If a customer seeks an adjustment because of defects in the merchandise, a markdown will be taken. If special sale merchandise is depleted during a sale, regular merchandise will be reduced in price to fill the demand. If wholesale cost is reduced during the season, price will be reduced correspondingly. If non-returnable merchandise is substandard on arrival, it will be reduced.

Most of the merchandise that becomes excess (especially for the major reasons outlined above) will be mentally transferred to an "availability pool." When a particular opportunity arises or when certain conditions develop necessitating a markdown, items are drawn out of this pool and marked down for the particular occasion involved. Store-wide clearances are planned by the merchandise manager on non-recurring dates throughout the year (except the pre-Fourth-of-July period and the after-Christmas period) to provide all departments an opportunity to clear out their excess stocks. The buyers receive a tentative schedule of these events ahead of time from the publicity department in a six-month's promotion plan bulletin.

However, there may be times during the year when the department

cannot wait for the next scheduled clearance for reasons of limited space capacity or of limited funds. If, for example, the department is expecting a large shipment of new merchandise and display and storage facilities do not have at the time the extra capacity to accommodate this additional inventory, there is no other choice but to clear some of the present stock by means of markdowns.

The department may take markdowns when its open-to-buy is too deep in the red. (The interpretation of "too deep" is somewhat arbitrary but whenever the open-to-buy falls to the minus \$15,000 level it would be judged to be in unsatisfactory condition.) The department will not necessarily take markdowns as the principal means of rectifying this state of affairs *per se* but instead will attempt to cancel merchandise on order or charge back merchandise already received *if* it cannot expect any relief from an increased sales rate within the immediate future or if the present average mark-up is low. But if a situation arises where the department has an urgent need to purchase additional merchandise and the open-to-buy at the time is in the red to the extent that the division merchandise manager will probably not approve any additional orders from the department, the department will then look to its availability pool and take markdowns for the amount necessary to permit the particular purchase to take place.

Amount of Markdown. The complete model for predicting actual markdown prices is given in Figure 5. The general rule for first markdowns is to reduce the retail price by $\frac{1}{3}$ and carry the result down to the nearest markdown ending (i.e., to the nearest \$0.85). There are some exceptions. Where the ending constraint forces too great a deviation from the $\frac{1}{3}$ rule (e.g., where regular price is \$5.00 or less), *ad hoc* procedures are occasionally adopted. On higher priced items, a 40% markdown is taken. On a few items manufacturers maintain price control. Occasionally, items represent a close out of a manufacturer's line and a greater markdown is taken.

Although the department did not seem to follow any particular explicit rule with second or greater markdowns, the higher the first markdown value the greater tended to be the reduction to the succeeding markdown price. In fact, this relationship seemed to follow the top half of the parabolic curve $Y^2 = 5(X - 2)$ where:

Y = Succeeding markdown price.

X = Initial markdown price.

Accordingly, the following empirically derived rule seems to work well with second or higher markdowns: "Insert the value of the initial markdown price in the parabolic formula and carry the result *down* to the nearest 85¢ (90¢)." As a description of process, this rule is obviously deficient. But in view of the limited number of cases involved and the inability of the department to articulate the rules, we have used this rough surrogate.

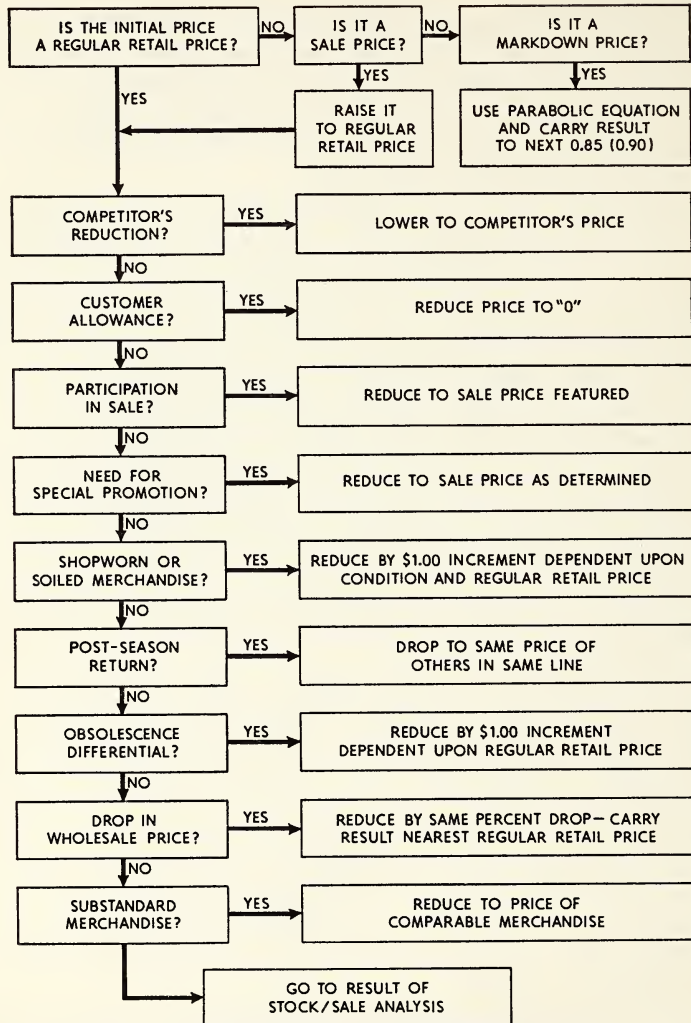


FIGURE 5. Beginning of flow chart for markdown routine.

TESTS OF THE PRICING AND ORDERING MODELS

We have tried to develop a model that would yield testable predictions. There are two major limits on such a goal. First, we have not been completely successful in defining a model that will make precise predictions in every decision area. Second, where we have been successful in developing a model, we are constrained by the availability of data. The value of data for the purpose of testing models has not always been controlling in data-retention decisions by the firm. Despite these limits, we have been able to develop models for the major price and ordering decisions and to subject all but one of the major components of those models to some empirical test.

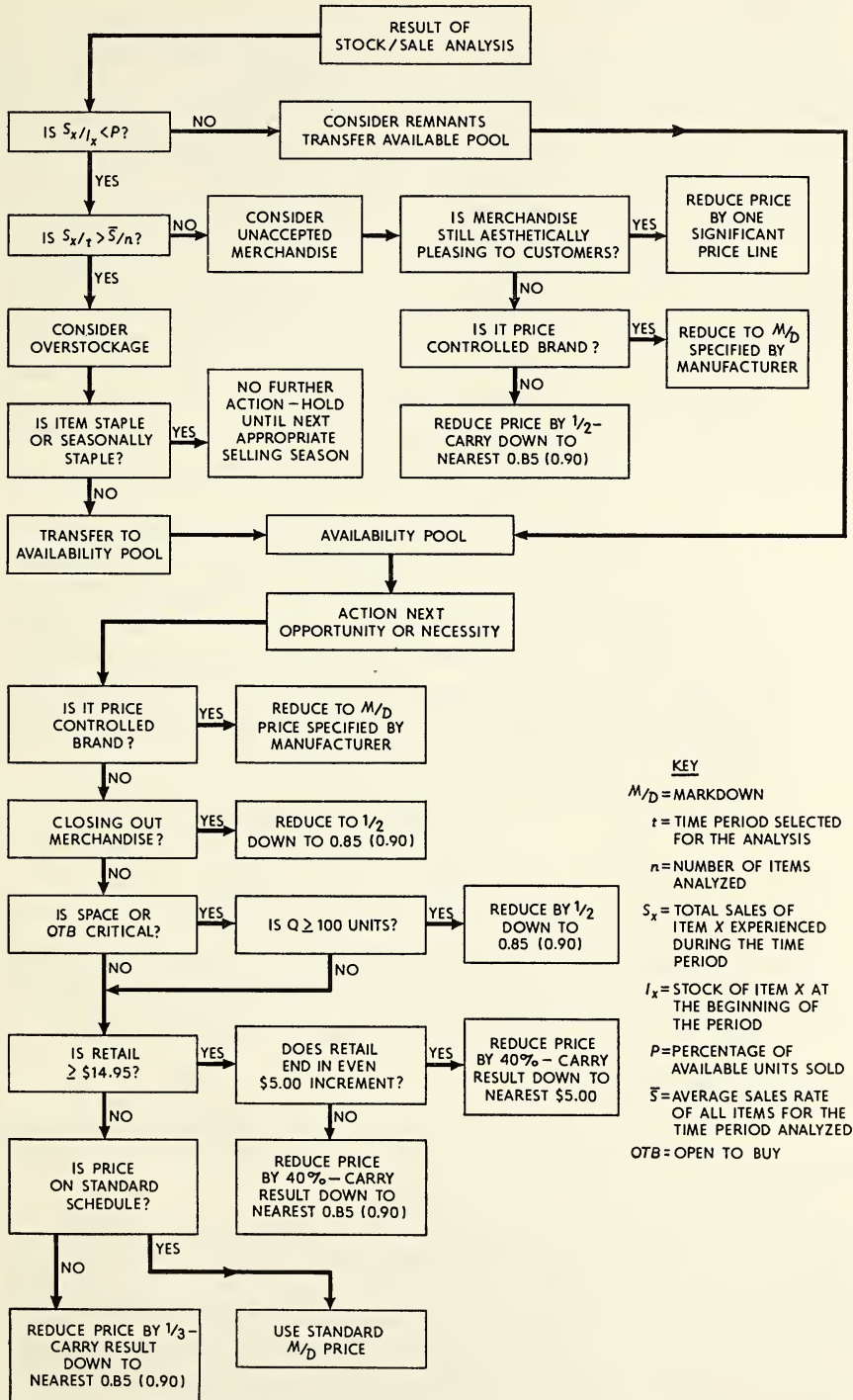


FIGURE 5 (Continued). Flow chart for markdown routine.

Determination of Orders

The ordering model consists of three segments—sales estimate, advance orders, and reorders. In each of the tests described below, the data used are new and are not the data with which the model was developed.

Sales Estimation. The sales estimation model is composed of the rule for estimating sales for the six-month period and of the rules for the estimation of the sales of individual months. The data available were for a two-year period so that the test is far from conclusive. However, there is no reason to believe that the model would not be valid for a larger sample of data. The first part of the model, the estimation of total sales for a six-month period, predicts the total within 5% in each of the four test instances. With the set of monthly rules, we can predict about 95% of the monthly sales estimates within 5%. There is no question that the predictive power could be increased still further by additional refinement of the rules. However, at this point it does not seem desirable to expend resources in that direction.

Advance Orders. This segment of the model and the sales estimation segment are related as we have shown previously. Therefore, discrepancies between predicted and actual are difficult to allocate precisely between the two segments although we have some clues from the above testing. Unfortunately the firm does not keep its records of advance orders any length of time so no extensive test of the model was possible. We were able to accumulate only four instances in which the predictions of the model could be compared with the actual.

Season	Predicted Advance Orders	Actual Advance Orders
1.....	18,050	16,453
2.....	26,550	24,278
3.....	36,200	35,922
4.....	43,000	35,648

Reorders. This segment is one that it is most important to test. The fact is, however, that the data on reorders are not kept in any systematic fashion, and we have not been able to make any kind of test.

Determination of Prices

The situation with regard to adequate data is much better for the pricing models. In each case the model was subjected to a large sample and performed adequately.

Mark-up. In order to test the ability of the model to predict the price decisions that will be made by the buyer on new merchandise, an unrestricted random sample of 197 invoices was drawn. The cost data and classification of the item were given as inputs to the computer model. The output was in the form of a predicted price. Since the sample consisted of items that had already been priced, it was possible to make a comparison of the predicted price with the actual.

The definition of a correct prediction was made as stringent as possible. Unless the predicted price matched the actual to the exact penny, the prediction was classified as incorrect. The results of the test were encouraging; of the 197 predicted prices 188 were correct and 9 were incorrect. Thus 95.4% of the predictions were correct. An investigation of the incorrect predictions showed that with minor modifications the model could be made to handle the deviant cases. However, at this point it was felt that the predictive power was good enough so that a further expenditure of resources in this direction was not justified.

Sale Pricing. In order to test the model a random sample of sales items was selected from the available records. A sample of 58 items was selected. For each item the appropriate information as determined by the model was used as an input to the computer. The output was in the form of a price which was a prediction of the price that would be set by the buyer. Again we used the criterion that to be correct the predicted price must match the actual price to the penny. Out of the 58 predictions made by the model, 56 were correct.

Markdowns. In testing this part of the model the basic data were taken from "markdown slips," the primary document of this firm. Naturally such slips do not show the information which would enable us to categorize the items for use in the model. It was necessary, therefore, to use direct methods such as the interrogation of the buyer and sales personnel to get the information necessary to classify the items so that the model could be tested. All of the data used were from the previous six-month period. It would be possible on a current basis to get the information which would enable the model to make the classifications itself as part of the pricing process.

The test for a correct prediction was the same as before, complete prediction, complete correspondence to the penny of the predicted and the actual price. A total sample of 159 items was selected and predictions made of the markdown price for each item. Of the 159 prices predicted, 140 were correct predictions by our criterion and 19 were wrong. This gives a record of 88% correct—the poorest record of the three models. Though this model does not do as well as the other two, the record is, in our view, adequate enough to allow reliance to be placed on the model.

SUMMARY

The tests that have been made of the model tend to support it. Clearly some of the tests are inadequate because of the paucity of the data. Also we have not attempted to build alternative models and compare predictive ability. No doubt that alternative models can be built and can be made to predict well. However, our primary interest has been in building a model which embodies the actual decision making process. We do not believe a radically different model can be built which captures the actual decision process. Because our objective is to understand the actual process, we have not attempted to mini-

mize the number of assumptions, the number of variables, or the number of inputs to the model.

The department store model outlined in this paper is supported by the ordering and pricing evidence collected in the field study. We would not argue that the evidence is conclusive with respect to the organization's decision making process. It is not. It is, however, consistent with the model. The model, in turn, lends itself to further elaboration and testing. And the world is full of firms for further empirical study.

A Heuristic Program for Locating Warehouses*

ALFRED A. KUEHN and MICHAEL J. HAMBURGER†

I. THE WAREHOUSE LOCATION PROBLEM

REGIONAL WAREHOUSES MAY PERFORM A VARIETY OF FUNCTIONS IN THE distribution of a manufacturer's products. These include: (1) the reduction of transportation costs relative to direct shipment to customers by permitting bulk or quantity shipments from factory to warehouse; (2) the reduction of delivery costs by combining products manufactured at several factories into single shipments to individual customers; and (3) the improvement of customer relations by decreasing delivery time relative to direct factory shipment, thereby permitting customers to reduce their inventories. There are, however, substantial costs associated with the operation of a regional warehouse system.

The problem at issue may therefore be phrased as follows: determine the geographical pattern of warehouse locations which will be most profitable to the company by equating the marginal cost of warehouse operation with the transportation cost savings and incremental profits resulting from more rapid delivery. A heuristic computer program which appears to be capable of generating reasonably good solutions to this class of problems will be described in the following sections, after a brief discussion of the heuristic approach to problem solving. A mathematical formulation of the warehouse location problem is given in Appendix I. A comparison of the heuristic program with several alternative approaches to the problem is contained in Appendix II.

*Reprinted by permission, with minor editorial revisions, from *Management Science*, Vol. IX. This research has been supported to varying degrees by the Graduate School of Industrial Administration and IBM and Ford Foundation Fellowships. While a number of individuals have offered valuable comments in reviews of earlier drafts of this paper, the authors particularly acknowledge the suggestions and encouragement of W. W. Cooper and Ralph L. Day of Carnegie Institute of Technology.

†Both of the Carnegie Institute of Technology.

II. THE HEURISTIC APPROACH TO PROBLEM SOLVING

Simon [27] has referred to heuristics as rules of thumb selected on the basis that they will aid in problem solving. In an earlier paper, Simon, in collaboration with Newell and Shaw, used the term "heuristic" to denote "any principle or device that contributes to the reduction in the average search to a solution" ([21], p. 22). Making use of the latter definition, a heuristic program can be defined (after Tonge [29]) as a problem-solving program organized around such principles or devices. Simon [27] has distinguished between such programs and algorithms on the basis that *only* the latter guarantee solution of the problem to a desired degree of accuracy. We do not believe that this is the most appropriate way to characterize heuristic programs. There are many solution procedures referred to as algorithms which do not guarantee solutions to a desired degree of accuracy, but rather, as is possible with the heuristic warehouse location program, provide only upper and lower bounds to the solution (for example, the fictitious play method for solving matrix games [17]). Furthermore the definition of algorithm generally used by mathematicians (for example, Courant and Robbins [8], p. 44) is "a systematic method for computation." Such a definition would include all computer programs.

We prefer to look at heuristic programming as an approach to problem solving where the emphasis is on working towards *optimum solution procedures* rather than *optimum solutions*. This is not to say that we ever expect to obtain an optimum solution procedure. The requirement of optimality would, in fact, be contradictory to the concept of using heuristic techniques. Heuristic techniques are most often used when the goal is to solve a problem so the solution is described in terms of acceptability characteristics rather than by optimizing rules (Tonge [29] p. 232).¹ The traditional operations research approach has been to search for optimum solutions. The heuristic approach differs in the following ways:

1. Explicit consideration is given to a number of factors (for example, computer storage capacity and solution time) in addition to the quality of the solution produced.
2. The evaluation of heuristic techniques is usually done by inductive rather than deductive procedures. That is, specific heuristics are justified not because they attain an analytically derived solution (for example, an optimum) but rather because experimentation has proved that they are useful in practice ([27], p. 11).

Recent interest in the heuristic approach to problem solving has led to the development of computer programs designed to: compose music [15], play checkers [24], play chess [5, 18, 20], discover proofs for

¹ These points have been discussed in more detail elsewhere in the context of a theory of human problem solving and choice (See, for example, Simon [26], pp. 196-207 and pp. 241-74, March and Simon [19], chaps. vi and vii, and Cyert, Dill, and March [9]).

theorems in logic and geometry [22, 12], design electric motors and transformers [14], and balance assembly lines [28].

III. A HEURISTIC PROGRAM FOR LOCATING WAREHOUSES

The heuristic program which we propose for locating warehouses consists of two parts: (1) *the main program*, which locates warehouses one at a time until no additional warehouses can be added to the distribution network without increasing total costs, and (2) *the bump and shift routine*, entered after processing in the main program is complete, which attempts to modify solutions arrived at in the main program by evaluating the profit implications of dropping individual warehouses or of shifting them from one location to another. The three principal heuristics used in the main program are:

1. *Most geographical locations are not promising sites for a regional warehouse; locations with promise will be at or near concentrations of demand.*²

The use of this heuristic in searching for and screening potential warehouse locations permits us to concentrate upon substantially less than $\frac{1}{100}$ of 1 percent of the United States and thereby eliminate mountains, marshes, deserts, and other desolate areas from consideration. To be sure, the program may as a result miss a good location. In general, however, computer time is put to much better use in screening and evaluating a finite number of concentrations of demand than in searching blindly for a possible profitable desolate location. (If management or the program operator is interested in evaluating any specific locations of this type, they can be entered as alternatives.)

2. *Near optimum warehousing systems can be developed by locating warehouses one at a time, adding at each stage of the analysis that warehouse which produces the greatest cost savings for the entire system.*

The use of this heuristic reduces the time and effort expended in evaluating patterns of warehouse sites. Thus, if there are M possible warehouse locations, the above heuristic would reduce the number of cost evaluations necessary from 2^M to approximately $N(M' + 1) < NM$ where N is the size of the intermediate buffer, discussed further below, and M' is the number of warehouses located. One can think of several classes of examples in which this heuristic would not work very well. However, such situations appear to occur only rarely in practice.

3. *Only a small subset of all possible warehouse locations need be evaluated in detail at each stage of the analysis to determine the next warehouse site to be added.*

To insure adding that warehouse location producing the greatest cost savings we could evaluate completely each of the remaining potential warehouse sites. The time required by such an approach can, however, be reduced very substantially with the addition of only slight risk with a good, easily computed method of screening potential sites. The heuristic used for screening calls for N of the M potential warehouse locations ($M > N > 1$) to be evaluated in detail at each stage (see step 3, Figure 1, Flow Diagram). The N potential warehouse sites chosen at each stage are those which, considering only local

² Baumol and Wolfe [4] also consider only a limited number of places at which to obtain warehouse space. However, their problem considers only the leasing of warehouse space and consequently the number of available locations is already limited and no heuristic is required to restrict the alternatives to be considered.

demand, would result in the greatest cost savings (or smallest increase in costs) if serviced by a local warehouse rather than by the system existing in the previous stage (see step 2, Flow Diagram). In other words, it is assumed that at any stage we can do reasonably well by locating the next warehouse in one of the N areas chosen on the basis of local demand and related warehousing and transportation costs.

In the detailed evaluation of each of the N locations placed in the buffer at each stage, the program either eliminates the site from further consideration, assigns a warehouse to that location, or returns the location to the list of potential warehouse sites for reconsideration at later stages in the program (steps 4, 6, and 7, respectively, in the Flow Diagram). Any site whose addition would not reduce total distribution costs is eliminated from further analysis in the main program. Of those sites which reduce total costs, that location which affords the greatest savings is assigned a warehouse; all others are returned to the list of potential warehouse sites. When the list of potential warehouses is depleted, all sites having been either eliminated or assigned a warehouse, the program enters the Bump and Shift Routine.

The Bump and Shift Routine is designed to modify solutions reached in the main program in two ways. It first eliminates (bumps) any warehouse which is no longer economical because some of the customers originally assigned to it are now serviced by warehouses located subsequently. Then, to insure the servicing of each of the territories established above from a single warehouse within each territory in the most economical manner, the program considers shifting each warehouse from its currently assigned location to the other potential sites (original list) within its territory. It should be noted that this routine does not guarantee that each territory will in fact be serviced in the most economical manner (this deficiency is illustrated below with reference to several sample problems).

The basic steps in the heuristic program are summarized in the Flow Diagram, Figure 1. Before going on to examine the results obtained in applying this program to several sample problems, let us discuss briefly the heuristics used in handling the shipping cost data (see step 1d, Flow Diagram). The inclusion of lists of actual transportation costs and delay times (or distance) between all potential shipping points as data might appear to be unwieldy relative to the Shycon-Maffei approach [25] since locating M customers and warehouses by longitude and latitude takes only M computer locations whereas up to $M^2/2$ locations are required for recording all shipping costs. In practice the problem is not nearly so unwieldy and produces some compensations: For example in many cases *a priori* judgments can be made that customers in certain geographical regions will not be serviced from potential warehouses in other regions. Thus, in one of the sample problems discussed below, we do not consider shipping from warehouses in Eastern cities to the West

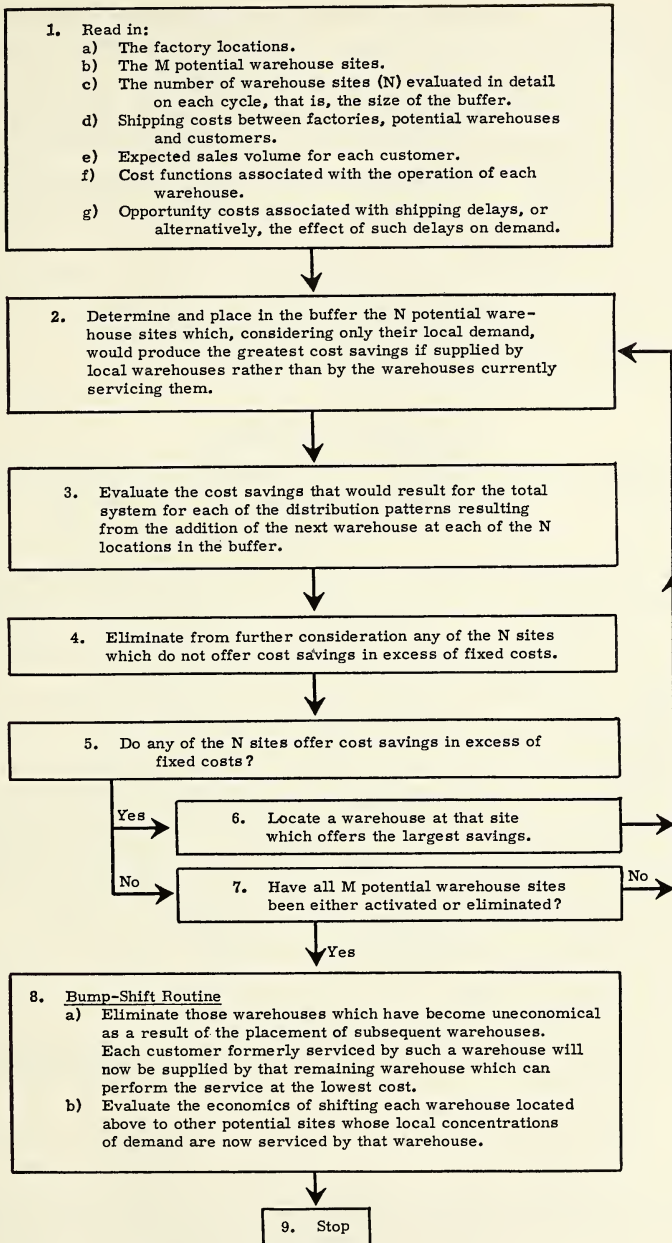


FIGURE 1. Flow diagram.

Coast since the factory is located in Indianapolis. In addition, customers can frequently be aggregated into concentrations of demand (for example, metropolitan chain grocery and wholesaler warehouses) because of close geographical proximity. The program also automatically

eliminates from the search list potential warehouse sites and warehouse-customer combinations that no longer offer promise of cost reduction. As a result, by the time that the program has located two or three warehouses, the list being searched is frequently reduced by 90 percent. The reduction in list size continues as more warehouses are added, speeding up the analysis on each cycle. These heuristics make the use of actual costs computationally efficient (computation times for 12 sample problems are outlined in the following section). They also permit us to: (1) avoid the errors associated with the use of air miles as a basis for approximating shipping costs; and (2) solve large-scale problems, involving, for example, several factories, 10 products, at least 200 potential warehouse sites, and more than a thousand concentrations of demand.

IV. SAMPLE PROBLEMS

The operation of the program will be illustrated with reference to 12 sample problems. The problems represent all combinations of three sets of factory locations—(1) Indianapolis, (2) Jacksonville, Florida, (3) Indianapolis and Baltimore—and four levels of fixed warehouse costs—\$7,500, \$12,500, \$17,500, \$25,000—for each warehouse in the system. Each of the sample problems considers only a single product. Transportation costs and costs associated with shipping delays are assumed to be proportional to the railroad distance between shipping points.³ For purposes of illustration, bulk shipping rates from the factory to warehouses are evaluated at \$0.0125 per mile per unit, whereas the sum of the shipping and delay costs from warehouses to customers is considered to be \$0.0250 per mile per unit. To further simplify the 12 distribution problems analyzed in this paper, the variable costs of operating the warehouses are assumed to be linear with respect to the volume of goods processed.⁴ Consequently, these costs do not affect the optimal warehouse system and need not be further considered in the sample problems.⁵ The size of

³ Railroad distances between cities were obtained from the *Rand-McNally Cosmopolitan World Atlas* (Chicago: Rand McNally & Co., 1951), p. 193. It should be noted that the simplifying assumption of linearity of transportation and delay costs with respect to railroad mileage is strictly a matter of convenience in making the cost data available to the reader. In practice, actual shipping costs and delay times would generally be read into the computer.

⁴ The 12 sample problems were not specified so as to fully test the generality of the program. The simplification of linear warehousing cost functions was incorporated in the test problems so that the heuristic solutions obtained might subsequently be compared with optimal solutions developed by the application of integer programming. Such a comparison has not yet been made since we are not aware of the existence of an integer programming routine capable of solving a problem of this size.

⁵ It should be noted that the heuristic program, in addition to being able to treat the case where both shipping costs and the variable and fixed costs of warehousing vary throughout the country, can determine which of several different types of warehouses (some of which might include packaging facilities) and transportation systems should be used to service each concentration of demand. The program can also be employed to locate regional factories, choosing among alternative factory sites for which production costs are specified.

the buffer (N) was equal to five in each of the 12 sample problems.

The market structure considered in the sample problems consists of 50 concentrations of demand scattered throughout the United States. Twenty-four of these centers of demand are treated as potential warehouse sites. The metropolitan population of each of these areas was used to represent sales potential, a population of 1,000 representing one unit of demand (see Table 4).

TABLE 1
HEURISTIC SOLUTIONS TO SAMPLE PROBLEMS
FACTORY LOCATION: INDIANAPOLIS
Fixed Costs of Warehouses

\$7,500		\$12,500		\$17,500		\$25,000	
<i>Warehouse Located at Each Stage</i>	<i>Cost of System at Each Stage</i>	<i>Warehouse Located at Each Stage</i>	<i>Cost of System at Each Stage</i>	<i>Warehouse Located at Each Stage</i>	<i>Cost of System at Each Stage</i>	<i>Warehouse Located at Each Stage</i>	<i>Cost of System at Each Stage</i>
<i>Main Program</i>		<i>Main Program</i>		<i>Main Program</i>		<i>Main Program</i>	
No warehouses...	\$1,248,688	No warehouses...	\$1,248,688	No warehouses...	\$1,248,688	No warehouses...	\$1,248,688
Philadelphia....	1,075,120	Philadelphia....	1,080,120	Philadelphia....	1,085,120	Philadelphia....	1,092,620
Los Angeles....	910,514	Los Angeles....	920,514	Los Angeles....	930,514	Los Angeles....	945,514
Seattle.....	876,429	Seattle.....	891,429	Seattle.....	906,429	Seattle.....	928,929
San Francisco...	861,967	San Francisco...	881,967	San Francisco...	901,967		
Houston.....	850,645	Houston.....	875,645	Houston.....	900,645	<i>Bump-Shift Routine</i>	
Chicago.....	839,853	Chicago.....	869,853	Chicago.....	899,853		
New York.....	830,424	New York.....	865,424			No change....	
Detroit.....	824,721	Detroit.....	864,721				
Denver.....	819,073	Kansas City...	860,484	<i>Bump-Shift Routine</i>		<i>Improvements Not Found by the Heuristic Program</i>	
Pittsburgh....	815,818	Atlanta.....	859,125	Replace Hous-			
Washington,		Cleveland.....	858,764	ton with		None known	
D.C.....	813,321	<i>Bump-Shift Routine</i>		Dallas.....	\$ 896,864		
Kansas City...	809,827	Drop Detroit...\$	857,725	<i>Improvements Not Found by the Heuristic Program</i>			
Boston.....	808,203	Replace Phila-					
Atlanta.....	801,845	delphia with Wash-		None known			
<i>Bump-Shift Routine</i>		ington.....	856,257				
Drop Denver...\$	801,748	<i>Improvements Not Found by the Heuristic Program</i>					
<i>Improvements Not Found by the Heuristic Program</i>		Replace Hous-					
Replace Hous-		ton with					
Dallas.....\$	800,163	Dallas.....\$	854,672				

The results obtained for each of the 12 cases are shown in Tables 1 through 3. These tables summarize:

1. The warehouse locations selected by the main program, in the order of selection.
2. The modifications introduced into the main program solution by the bump-shift routine.
3. Alterations to the heuristic warehouse network which are known to lower total distribution costs.

4. The total distribution costs at each stage of the heuristic solution and for the warehouse network which incorporates subsequent improvements.⁶

In each of the four cases in which an improvement upon the heuristic solution was discovered, the improvement consisted of replacing a warehouse in Houston with a warehouse in Dallas. This improvement was not found by the shift portion of the bump-shift routine since Dallas

TABLE 2
HEURISTIC SOLUTIONS TO SAMPLE PROBLEMS
FACTORY LOCATION: JACKSONVILLE
Fixed Costs of Warehouses

\$7,500		\$12,500		\$17,500		\$25,000	
<i>Warehouse Located at Each Stage</i>	<i>Cost of System at Each Stage</i>	<i>Warehouse Located at Each Stage</i>	<i>Cost of System at Each Stage</i>	<i>Warehouse Located at Each Stage</i>	<i>Cost of System at Each Stage</i>	<i>Warehouse Located at Each Stage</i>	<i>Cost of System at Each Stage</i>
<i>Main Program</i>		<i>Main Program</i>		<i>Main Program</i>		<i>Main Program</i>	
No warehouses.....	\$1,832,861	No warehouses.....	\$1,832,861	No warehouses.....	\$1,832,861	No warehouses.....	\$1,832,861
New York.....	1,602,504	New York.....	1,607,504	New York.....	1,612,504	New York.....	1,620,004
Los Angeles.....	1,376,900	Los Angeles.....	1,386,900	Los Angeles.....	1,396,900	Los Angeles.....	1,411,900
Chicago.....	1,239,864	Chicago.....	1,254,864	Chicago.....	1,269,864	Chicago.....	1,292,364
Seattle.....	1,211,687	Seattle.....	1,231,687	Seattle.....	1,251,687	Seattle.....	1,281,687
Washington, D.C.....	1,169,916	Washington, D.C.....	1,194,916	Washington, D.C.....	1,219,916	Washington, D.C.....	1,257,416
St. Louis.....	1,143,998	St. Louis.....	1,173,998	St. Louis.....	1,203,998	St. Louis.....	1,248,998
Cincinnati.....	1,123,224	Cincinnati.....	1,158,224	Cincinnati.....	1,193,224	Cincinnati.....	1,245,724
Houston.....	1,106,239	Houston.....	1,146,239	Houston.....	1,136,239		
San Francisco.....	1,098,949	San Francisco.....	1,143,949				
Denver.....	1,096,245			<i>Bump-Shift Routine</i>		<i>Bump-Shift Routine</i>	
Detroit.....	1,093,780	<i>Bump-Shift Routine</i>		No change.....	\$1,136,239	No change.....	\$1,245,724
Pittsburgh.....	1,092,163	No change.....\$1,143,949		<i>Improvements Not Found by the Heuristic Program</i>		<i>Improvements Not Found by the Heuristic Program</i>	
Atlanta.....	1,091,374	<i>Improvements Not Found by the Heuristic Program</i>		None known		None known	
<i>Bump-Shift Routine</i>		<i>Improvements Not Found by the Heuristic Program</i>					
No change.....	\$1,091,374	None known					
<i>Improvements Not Found by the Heuristic Program</i>							
None known							

⁶The improvements upon the heuristic solutions tabulated for four of the 12 problems have been found by evaluating the modifications in the heuristic distribution network which, upon inspection, appeared most likely to result in lower distribution costs. Approximately 25 types of modifications were tested, but only one resulted in a minor improvement in four of the problems. To determine the optimal system by complete enumeration would require the evaluation of the 2²⁴ possible ways of locating up to 24 warehouses. With an IBM-704, which could perform approximately one such evaluation per second, this operation would take more than six months of continuous operation per problem. Testing all combinations of three and four warehouses, which would probably be sufficient to insure finding the optimal network for any one of the two least interesting cases (those problems in which, because of high fixed warehousing costs, the heuristic program locates only three warehouses) would require approximately four computer hours. The use of integer programming offers more promise. Once such a computer program is available, it should be feasible to test the sample problem solutions we have found for optimality in one iteration by using the heuristic solutions as the initial basis for the integer

TABLE 3
HEURISTIC SOLUTIONS TO SAMPLE PROBLEMS
FACTORY LOCATIONS: BALTIMORE AND INDIANAPOLIS
Fixed Costs of Warehouses

\$7,500		\$12,500		\$17,500		\$25,000	
<i>Warehouse Located at Each Stage</i>	<i>Cost of System at Each Stage</i>	<i>Warehouse Located at Each Stage</i>	<i>Cost of System at Each Stage</i>	<i>Warehouse Located at Each Stage</i>	<i>Cost of System at Each Stage</i>	<i>Warehouse Located at Each Stage</i>	<i>Cost of System at Each Stage</i>
<i>Main Program</i>		<i>Main Program</i>		<i>Main Program</i>		<i>Main Program</i>	
No warehouses . . .	\$899,770	No warehouses . . .	\$899,770	No warehouses . . .	\$899,770	No warehouses . . .	\$899,770
Los Angeles	735,164	Los Angeles	740,164	Los Angeles	745,164	Los Angeles	752,664
Seattle	701,079	Seattle	711,079	Seattle	721,079	Seattle	736,079
New York	672,666	New York	687,666	New York	702,666	New York	725,166
San Francisco	658,205	San Francisco	678,205	San Francisco	698,205		
Houston	646,882	Houston	671,882	Dallas	691,772	<i>Bump-Shift Routine</i>	
Chicago	636,091	Chicago	666,091	Chicago	690,980	No change	\$725,166
Denver	630,442	Kansas City	661,853	<i>Bump-Shift Routine</i>		<i>Improvements Not Found by the Heuristic Program</i>	
Detroit	626,481	<i>Bump-Shift Routine</i>		<i>Improvements Not Found by the Heuristic Program</i>		<i>Improvements Not Found by the Heuristic Program</i>	
Kansas City	623,114	<i>Improvements Not Found by the Heuristic Program</i>		<i>Improvements Not Found by the Heuristic Program</i>		<i>Improvements Not Found by the Heuristic Program</i>	
Cleveland	620,405	<i>None known</i>		<i>None known</i>		<i>None known</i>	
Atlanta	617,337	No change	\$661,853	<i>None known</i>		<i>None known</i>	
<i>Bump-Shift Routine</i>		<i>Improvements Not Found by the Heuristic Program</i>		<i>Improvements Not Found by the Heuristic Program</i>		<i>Improvements Not Found by the Heuristic Program</i>	
Drop Denver	\$617,116	Replace Hous- ton with Dallas	\$660,268	<i>None known</i>		<i>None known</i>	
<i>Improvements Not Found by the Heuristic Program</i>		<i>None known</i>		<i>None known</i>		<i>None known</i>	
Replace Hous- ton with Dallas	\$615,531	<i>None known</i>		<i>None known</i>		<i>None known</i>	

TABLE 4
SALES POTENTIAL OF CONCENTRATIONS OF DEMAND
USED IN SAMPLE PROBLEMS
(Population in Thousands)

<i>Concentrations of Demand</i>	<i>Sales Potential</i>	<i>Concentrations of Demand</i>	<i>Sales Potential</i>
Albuquerque, N.Mex.....	146	Knoxville, Tenn.....	337
Amarillo, Texas.....	87	Los Angeles, Calif.*.....	4,368
Atlanta, Ga.*.....	672	Louisville, Ky.....	577
Baltimore, Md.....	1,337	Memphis, Tenn.*.....	482
Billings, Mont.....	31	Miami, Fla.*.....	495
Birmingham, Ala.....	559	Mobile, Ala.....	231
Boston, Mass.*.....	2,370	Nashville, Tenn.....	322
Buffalo, N.Y.*.....	1,089	New Orleans, La.*.....	685
Butte, Mont.....	33	New York, N.Y.*.....	12,912
Cheyenne, Wyo.....	32	Oklahoma City, Okla.....	325
Chicago, Ill.*.....	5,495	Omaha, Nebr.....	366
Cincinnati, Ohio*.....	904	Philadelphia, Pa.*.....	3,671
Cleveland, Ohio*.....	1,466	Pittsburgh, Pa.*.....	2,213
Columbia, S.C.....	143	Portland, Oregon.....	705
Dallas, Texas*.....	615	Richmond, Va.....	328
Denver, Colo.*.....	564	St. Louis, Mo.*.....	1,681
Des Moines, Iowa.....	226	St. Paul, Minn.*.....	1,117
Detroit, Mich.*.....	3,016	Salt Lake City, Utah*.....	275
Duluth, Minn.....	253	San Antonio, Texas.....	500
El Paso, Texas.....	195	San Francisco, Calif.*.....	2,241
Fargo, N.Dak.....	38	Seattle, Wash.*.....	733
Houston, Texas*.....	807	Spokane, Wash.....	222
Indianapolis, Ind.....	551	Tucson, Ariz.....	49
Jacksonville, Fla.....	304	Washington, D.C.*.....	1,464
Kansas City, Mo.*.....	814	Wichita, Kansas.....	222

* Potential warehouse sites.

Source: *The World Almanac* (New York: *New York World-Telegram and the Sun*, 1960).

neighboring warehouse sites of each potential warehouse location are specified in the input to the program and are evaluated as alternatives in the shift routine. Since this routine does not make use of the concept of warehouse territories, it will also correct errors of the Dallas-Houston variety when applied to problems involving only a single product. (This routine will increase computation time in most cases.) It will not, however, correct for less-localized deviations of the main program solution from the optimal warehouse network.

In considering the development of new or more elaborate bump-shift routines (or, the inclusion of such devices in the main program to be performed after the addition of each warehouse) it is desirable to determine the improvements which might be expected. In the 12 sample problems discussed above, the improvements upon the main program solution developed in the bump-shift routine and through subsequent analysis never amounted to more than 0.5 percent. If future research indicates that the main program solution is generally near-optimal, it would give strong support to the use of the three basic heuristics in the main pro-

gram and limit the gain to be expected in searching for corrective routines.

Some support for the heuristic which selects warehouse sites to be placed in the buffer is obtained by reference to the frequency distribution of warehouses selected for activation in the 12 sample problems from each of the positions in the five-place buffer used in the analysis. This distribution is tabulated in Table 5, the buffer positions representing the rank of the potential warehouse sites in terms of their cost savings considering only local demand (Step 2, Figure 1, Flow Diagram).

TABLE 5
FREQUENCY DISTRIBUTION OF WAREHOUSES SELECTED
FOR ACTIVATION FROM EACH POSITION IN THE BUFFER

<i>Position in the Buffer</i>	<i>Number of Warehouses Located from Each Position</i>	<i>Percentage of Total Warehouses Located from Each Position</i>
1.....	48	49.0%
2.....	15	15.3
3.....	15	15.3
4.....	10	10.2
5.....	10	10.2

Computation Time

The time required to reach a solution for the 12 sample problems in the main program totaled 72 minutes on an IBM-650 with RAMAC disc storage. The individual problems required an average of 2 minutes setup time and 30 seconds per warehouse located. Experimentation with and analysis of the heuristic program indicates that computation time increases at a much slower rate with increases in problem size than is the case with linear programming algorithms designed to handle fixed cost elements. It appears that the problem setup time increases linearly with the product of the number of warehouses, the number of products, and the number of customers (concentrations of demand). The time required for locating warehouses increases approximately linearly with the size of the buffer (N), the number of products, and the number of customers, but almost negligibly with the number of potential warehouse sites. The effect of multiple factories on setup time is at most linear; if capacity constraints are not operative the effect is substantially less than linear. Surprisingly, increasing the number of factories actually tends to decrease the total warehouse location time since there is no effect upon the time required to locate individual warehouses and the total number of warehouses located will generally be reduced.

Accurate time estimates for the bump-shift routine are not available since only an inefficient routine was operating when the IBM-650 at Carnegie Tech was replaced. Processing the 12 sample problems with this version of the bump-shift routine required a total of one hour. We

would expect an efficient computer routine to perform this operation in 10 to 15 minutes since comparable reductions in computer time were achieved in the revision of the main program.

Extensions and Applications

Improved heuristics in terms of reduced computation time and/or more nearly optimum solutions of the warehouse location problem will probably be forthcoming. The six-minute solution time on an IBM-650 in itself probably suggests to the reader that additional check and bump-shift routines might be interspersed between the location of individual warehouses; after all, the difference in cost between six minutes and even six hours of computation time on an IBM-650 is negligible relative to the cost savings that might be achieved in the sample warehouse network problems studied. It is not clear, however, that such approaches will on the average improve upon the solutions generated by the existing heuristic program if at least four warehouses are located. Care must also be taken to avoid the chase for optimal solutions to simple problems and thereby miss the actual problem of business—the solution of large-scale problems containing many customers buying various mixes of a full product line, many potential warehouse sites, alternate warehouse types with different cost structures, several factories and, perhaps, a number of potential factory sites.⁷

Two warehouse network problems containing many of the above complexities are now being examined, one representing the distribution of a variety of grocery-drug type consumer products, the other a line of consumer appliances. In both cases the problem is of such magnitude that the solution time could easily come to several hours on an IBM-7090 computer unless the problem is simplified beyond the level now thought to be desirable. Furthermore, in view of some uncertainty as to the actual nature of warehousing costs, it appears prudent to make several runs with different warehousing cost functions to determine the sensitivity

⁷ It should be noted that the distribution of order-shipment mixes of products can be treated by considering each geographical concentration of demand as several concentrations at the same location, each concentration representing a given mix of products of a given total size. Insofar as discrete distributions can be used to approximate the empirical distributions, the specification of the problem can be greatly simplified. The number of computer locations required to store each of these mixes would then be reduced substantially from that which would be required if each of the mixes were treated as the demand at an individual geographical location. An interesting aspect of this treatment of the problem is that the total warehouse network would be established with full recognition of the fact that customers will not necessarily receive all of their shipments from a single warehouse if all factories do not produce all products. An order of packaged detergents and toilet bar soap, for example, might be received from one warehouse if it consists largely of detergent and from another if the order is primarily composed of bar soaps. Similarly, not all products will necessarily be stocked at all warehouses. In the distribution of appliances, for example, yellow refrigerators, for which there is a relatively small demand, might be stocked in only the larger warehouses.

of the heuristic warehouse solutions to the cost functions. It is in this context that the advantages of improved solutions to test problems must be evaluated relative to increased computer time requirements. Additional check or bump-shift heuristics increasing computer time on test problems from six minutes to six hours would increase the IBM-7090 time per run from 3-6 hours to 180-360 hours, time which might well be better spent in testing the sensitivity of the solutions to variations in warehouse cost structures and in providing an improved model of demand through greater detail in the description of the product mix and size of customer orders.

It has been suggested that the heuristics might be applied by eliminating warehouses rather than adding them. That is, a warehouse is assumed to be operating at each potential site at the start of the program. Warehouses would then be eliminated one by one on the basis of cost savings. It is not clear how the quality of solutions produced in this manner would compare with those developed from the heuristic program outlined in this paper. In terms of computation time, however, it seems likely that the current program would be the more efficient when the number of warehouses located is less than half the number of potential sites being considered. This seems to be generally the case in industry although situations could, no doubt, be found where this is not true (for example, the firm which is interested only in considering the possibility of closing existing warehouses).

Once we know the optimal solutions to the test problems (through the application of integer programming, using the heuristic solution as a starting basis), we will be in a better position to evaluate the potential gains possible through the use of improved heuristics. In addition, knowledge of the optimal solution should provide sound direction as to the types of heuristics which offer most promise in correcting the deviations inherent in the current program.

V. SUMMARY

A heuristic program was developed and applied to several warehouse location problems. The results suggest that a heuristic approach to this class of problems may be quite profitable in practice, producing near-optimal solutions within acceptable limits of computer time.

The use of heuristics in solving these problems has two prime advantages relative to the currently available linear programming formulations and solution procedures: (1) computational simplicity, which results in substantial reductions in solution times and permits the treatment of large-scale problems, and (2) flexibility with respect to the underlying cost functions, eliminating the need for restrictive assumptions. It also offers an important advantage relative to the simulation technique of Shycon and Maffei in that it incorporates a systematic procedure designed to generate at least one near-optimal distribution system while providing approximately the same flexibility in the modeling of the problem.

The proposed heuristic program permits fast screening and evaluation of

alternative types of warehouses, transportation systems, and warehouse locations. It should, however, be emphasized that this program is not the end of the road. It may some day become practical to solve large-scale warehouse location problems with optimizing algorithms given continued development of computer hardware and linear programming techniques. Heuristic programming, too, is capable of improvement and such developments will probably be forthcoming as a result of the large amount of research on heuristic models and computer programming now being carried on at the RAND Corporation, Carnegie Tech, and elsewhere.

APPENDIX I. MATHEMATICAL FORMULATION OF THE WAREHOUSE LOCATION PROBLEM

The problem can be expressed mathematically as follows:

- $X_{h,i,j,k}$ = the quantity of good h ($h = 1, \dots, p$) shipped from factory i ($i = 1, \dots, q$) via warehouse j ($j = 1, \dots, r$) to customer k ($k = 1, \dots, s$).
- $A_{h,i,j}$ = the per unit transportation cost of shipping good h from factory i to warehouse j .
- $B_{h,j,k}$ = the per unit transportation cost of shipping good h from warehouse j to customer k .
- $C_{h,j}(\sum_{i,k} X_{h,i,j,k})$ = total cost of warehouse operation associated with processing good h at warehouse j . Without loss of generality we may express this function as the sum of $S_{h,j}$ and F_j defined below.
- $D_{h,k}(T_{h,k})$ = explicit or imputed cost due to a delay of T time units in delivery of good h to customer k . When the customer imposes a maximum delivery time (constraint), D becomes infinite whenever the indicated limit is reached.⁸
- F_j = fixed cost per time period of operating warehouse j . Note that this is a planned fixed cost to be incurred and not a sunk cost.
- $S_{h,j}(\sum_{i,k} X_{h,i,j,k})$ = semivariable cost of operating warehouse j per unit of good h processed, including variable handling and administrative costs, storage costs, taxes, interest on investment, pilferage, and so on (the homogeneous portion of the very general function $C_{h,j}$).
- $Q_{h,k}$ = quantity of good h demanded by customer k .
- W_j = capacity of warehouse j .
- $Y_{h,i}$ = capacity of factory i to produce good h .
- $Z_j = 1$ if $\sum_{h,i,k} X_{h,i,j,k} > 0$ and zero otherwise (that is, $\sum Z_j$ = the number of warehouses used).

The problem then becomes one of minimizing total distribution costs, an objective function of the form

$$f(X) = \sum_{h,i,j,k} (A_{h,i,j} + B_{h,j,k}) X_{h,i,j,k} + \sum_j F_j Z_j + \sum_{h,j} S_{h,j}(\sum_{i,k} X_{h,i,j,k}) + \sum_{h,k} D_{h,k}(T_{h,k}),$$

subject to constraints of the following form:

$$\sum_{i,j} X_{h,i,j,k} = Q_{h,k}$$

(customer k 's demand for product h must be supplied),

⁸The effect of delay in supplying customers has been treated above as an "opportunity cost" since this simplifies the notation and is consistent with current research practice. However, an alternative formulation which reflects management's view of the problem more accurately is to have delivery times affect demand.

$$\sum_{j,k} X_{h,i,j,k} \leq Y_{h,i}$$

(factory i 's capacity limit on good h cannot be exceeded),

$$I_j (\sum_{h,i,k} X_{h,i,j,k}) \leq W_j$$

(the capacity of warehouse j cannot be exceeded),

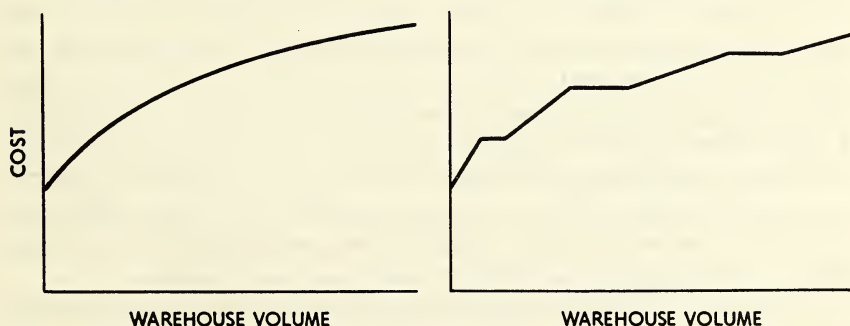
where $I_j (\sum_{h,i,k} X_{h,i,j,k})$ is a function which denotes the maximum inventory level associated with the flow of all goods from all factories to all customers serviced through warehouse j .

APPENDIX II. A COMPARATIVE STUDY OF ALTERNATIVE APPROACHES TO THE WAREHOUSE LOCATION PROBLEM

Linear Programing

In theory a linear programing approach (for example, using Gomory's integer algorithm [13] or other variations) could be used to solve the problem. In practice, however, the size and nonlinearities involved in many problems are such that application is not currently feasible.⁹

The most important types of nonlinearities generally encountered stem from the fixed costs associated with the operation of warehouses and variable warehousing and delivery costs which are nonlinear. Two types of warehousing and delivery cost functions have been suggested in the literature, strictly concave functions and piecewise linear functions. Figure 2 describes the strictly concave cost function proposed by Baumol and Wolfe [4]. Figure 3 represents the piecewise linear function which, apart from the fixed cost element, is equivalent to that illustrated by Balinski and Mills [2]. The fixed cost element is incorporated in Figure 3 for purposes of generality.¹⁰



Functions describing warehousing and delivery costs: FIGURE 2. Strictly concave function (left). FIGURE 3. Piecewise linear function (right).

⁹ The difficulties that have been encountered in applying linear programming directly to the warehouse location problem are similar to those encountered in the application of these algorithms to production scheduling problems. See for example, C. C. Holt, F. Modigliani, J. F. Muth, and H. A. Simon [16], chap. xx.

¹⁰ In the previous formulation of the warehouse location problem the warehouse operating costs and the cost of delivering goods to warehouses were separated. However, these two types of costs may be combined into one function without any loss of generality.

Both of the above functions describe essentially the same types of cost structures, that is, cost structures where the transportation rates and the marginal cost of operating a warehouse decrease as the quantity of goods handled by the warehouse increases. The different functions reflect slightly different views of the total problem, Baumol and Wolfe focusing primarily upon the warehouse operating costs which, in a sample problem, they approximate by square root functions (see [4], p. 255) and Balinski and Mills concentrating upon freight and storage rates specified in terms of the volume of goods handled.

Let us now examine several other approaches to the warehouse location problem discussed in the literature, paying particular attention to the treatment of those aspects of the problem not handled adequately by linear programming in its current state of development.

The Baumol-Wolfe Marginal Cost Approach

Baumol and Wolfe [3, 4] pose the problem as minimizing total distribution costs which, utilizing the symbols defined above, may be stated as follows:

$$\text{Min. } f(X) = \text{Min. } \sum_{i,j,k} (A_{ij} + B_{j,k}) X_{i,j,k} + \sum_j S_j (\sum_{i,k} X_{i,j,k}) + \sum_j F_j Z_j$$

subject to

$$I_j (\sum_{i,k} X_{i,j,k}) \leq W_j$$

(the warehouse capacity constraint), and

$$\sum_{i,j} X_{i,j,k} = Q_k$$

(customer demand constraint, that is, all demands must be satisfied).

The effects of delay in delivery times are not incorporated into the Baumol-Wolfe treatment of the problem although it is apparent that opportunity costs associated with such delays could be added to the shipping costs. It should also be noted that the above symbolic representation of the problem requires only three subscripts. The reason for this is that the Baumol and Wolfe method deals with only one product or, perhaps, a composite product (that is, a constant product mix). It is possible that the method can be extended to deal with the multiproduct case although Baumol and Wolfe do not indicate how this might be done.

The Baumol-Wolfe algorithm consists of an iterative procedure which requires the solution of an ordinary linear programming problem at each stage. The steps involved in this procedure are as follows:

1. *The Initial Stage.* For each pair i, k of factory and retailer, find the least cost of shipment considering *only* the transportation costs:

$$C_{i,k}^0 = \text{Min.}_j (A_{i,j} + B_{j,k}) = (A_{i,j_{i,k}}^0 + B_{j_{i,k},k})$$

where $j_{i,k}$ denotes the routing between factory i and customer k via warehouse j as selected by this criterion. The superscripts refer to the stage in the solution process. Thus, $C_{i,k}^0$ is the lowest unit cost of shipping

the product from factory i to customer k considered over all warehouse possibilities at the initial stage. These $C_{i,k}^0$ are then used to solve an ordinary transportation problem involving shipments from factories with known availabilities to retailers with known demands so as to minimize the cost function $\sum_{i,k} C_{i,k}^0 X_{i,k}$.¹¹

2. *The n^{th} Stage.* Thus starting at stage 0, the iterations are continued until a stage, say $n - 1$, is reached. From the warehouse loadings computed at stage $n - 1$, $(\sum_{i,k} X_{i,j,k}^{n-1})$, the marginal warehousing costs are computed by means of the expression

$$\frac{dS_j(\sum_{i,k} X_{i,j,k}^{n-1})}{d(\sum_{i,k} X_{i,j,k}^{n-1})},$$

the derivative of the operating costs of warehouse j .

The new set of transportation costs to be used in the transportation problem at this stage is defined as

$$C_{i,k}^n = \text{Min}_j A_{i,j} + B_{j,k} + \frac{dS_j(\sum_{i,k} X_{i,j,k}^{n-1})}{d(\sum_{i,k} X_{i,j,k}^{n-1})}$$

3. *Test for Stopping.* Compare the warehouse loadings of the n^{th} stage with those of the previous stage. If the same stop, if different continue by returning to step 2.

The justification for this approach is contained in a set of proofs which Baumol and Wolfe present in their paper. They do not claim that the method will produce an optimal solution but rather that

- a) "this procedure will reduce the total cost . . . in each step."¹²
- b) "the procedure . . . will yield a solution that probably cannot be improved except by prohibitively expensive computation for large-scale problems."¹³ That is, the method determines a local optimum.¹⁴

The Baumol-Wolfe method may be satisfactory for solving certain special types of warehouse location problems. However as a general approach to this class of problems the method appears to be unsatisfactory. First, although the formulation of the problem permits fixed costs associated with the operation of warehouses, there appears to be no mechanism within the solution procedure itself designed to handle this aspect of the problem. That is, both the differential calculus as used in adjusting the transportation costs and the transportation model are not well designed to handle fixed cost elements. Furthermore, even if fixed costs are neglected, the solutions to concave programming problems¹⁵ pro-

¹¹ Many efficient computing methods for solving this special form of linear programming are available [10, 6, 11].

¹² Baumol and Wolfe [4], p. 257.

¹³ *Ibid.*, p. 256.

¹⁴ Baumol [3], p. 413, and Baumol and Wolfe [4], p. 252.

¹⁵ That is problems in which the cost function possesses the property that $\frac{dC(x)}{dx} < \frac{C(x)}{x}$. For a more precise definition see Charnes and Cooper [7], p. 284.

duced by the Baumol-Wolfe method will frequently not be satisfactory. The quality of the solutions produced will in general depend on the degree of concavity of the cost function. To illustrate the impact which a concave cost function can have upon the solution to even a simple problem, let us consider the example used by Baumol and Wolfe to demonstrate their procedure.

The Baumol-Wolfe sample problem involves two factories, five potential warehouse sites and eight retailers. The problem involves no fixed costs in the operation of warehouses, warehousing costs being regarded as square root functions of warehouse volume. Table 1 shows the solutions derived by several procedures.

TABLE 6

<i>Method of Solution</i>	<i>Warehouses Used</i>	<i>Cost</i>
Baumol-Wolfe.....	1, 2, 4, 5	\$2,362*
Heuristic program.....	3	\$2,047
Arbitrary choice of single warehouses..	1	\$2,452
	2	\$2,379
	4	\$2,194
	5	\$2,070

*Baumol and Wolfe [4] calculated the total cost associated with their warehouse network solution as \$2,257, which is \$105 less than that tabulated above. This difference is due to an error in the computation of factory to warehouse costs. It should be noted, however, that this error does not in any way influence the Baumol-Wolfe solution with respect to warehouse locations or loadings.

The above table shows that the Baumol-Wolfe warehouse location procedure has a strong bias in the direction of locating more than the optimal number of warehouses. Baumol and Wolfe recognized the existence of such a bias ([4], p. 262), but did not, in their paper, concern themselves with the extent of its impact. To be sure, the Baumol-Wolfe solution is not even a local optimum in terms of changes in warehouse location patterns; a reduction in total costs can be accomplished by eliminating any one of the four warehouses in the solution set.¹⁶ The heuristic solution, obtained through application of the program discussed in this paper, has since been proven to be the optimal solution.

Balinski-Mills Average Cost Method

The Balinski-Mills approach to the warehouse location problem is similar to that of Baumol and Wolfe in that it requires the problem to be cast in the framework of the linear programming transportation model [2]. However, that is as far as the similarity goes. Baumol and Wolfe

¹⁶The Baumol-Wolfe solution is a local optimum in the sense that no individual unit of product can be shipped by an alternate route without increasing total distribution costs. Costs can be reduced, however, by transferring the total demand of a retailer (or a major portion thereof) from one warehouse to another or, as indicated above, by eliminating individual warehouses. Baumol [3] did apparently believe that his solution was a local optimum in the latter sense when he wrote, "No *minor* changes from this computed set of warehouse locations could yield any reduction in costs (the result was sure to be a local optimum)."

viewed the costs of delivering goods to warehouses and storing them as being described by a concave function whereas Balinski and Mills treated these costs as a piecewise linear function. Furthermore, Balinski and Mills specifically restrict their solution procedure to the case of a single factory (or vendor) and a single product ([2] p. 1).

In terms of the symbols defined above, Balinski and Mills pose the following problem:

$$\text{Minimize } \sum_j C_j(\sum_k X_{j,k}) + \sum_{j,k} B_{j,k} X_{j,k}$$

(total distribution costs)

$$\text{subject to } I_j(\sum_k X_{j,k}) \leq W_j$$

(the warehouse capacity constraint), and

$$\sum_j X_{j,k} = Q_k$$

(the customer demand constraint),

where the variable portion of $C_j(\sum_k X_{j,k})$ is a piecewise linear function (see Figure 3).

In principle, an optimal solution to this problem can be found by using integer programming. However as Balinski and Mills point out ([2] p. 6) even small sample problems often involve so many variables and constraints that they can not now be handled on existing computing equipment. As a result of these restrictions on the current use of integer programming, they develop an approximation technique which they report has proven to be highly successful in certain cases.¹⁷ This technique consists of approximating

$$C_j(\sum_k X_{j,k}) \text{ by } C_j(a_j)/a_j,$$

where $a_j = \min (W_j, \sum_{j,k} X_{j,k})$. That is, the warehousing cost function for each warehouse is approximated by the average unit cost of operating that warehouse at some high level, such as the warehouse capacity (W_j) or the total flow of goods passing through the system during the year.

By using the above approximation of the warehousing cost function all nonlinearities are eliminated from the problem. As a result the problem is cast into the framework of a simple transportation problem. The solution to this transportation problem yields a distribution network which minimizes the objective function

$$f(X) = \sum_{j,k} (B_{j,k} + C_j(a_j)/a_j) X_{j,k}$$

subject to the constraints listed above. The value of this solution, denoted by λ_* represents a lower bound to the cost of the optimal solution to the original problem. The $X_{j,k}$ shipment routings arrived at through the use of this algorithm are regarded as the solution to the original problem. The value (cost) of this solution to the original problem designated by λ^* , can then be calculated by substituting these $X_{j,k}$

¹⁷ Balinski and Mills [2], p. 6.

into the original cost functions. Balinski and Mills prove ([2] p. 8) that the cost of the actual optimal solution to the original problem (denoted by λ^0) must lie between these two values, that is, $\lambda_* < \lambda^0 < \lambda^*$.

Thus we see that the Balinski-Mills method possesses the very desirable property of providing a lower bound to the optimal solution λ^0 to help evaluate the quality of the solution λ^* . Such a property is desirable since it cautions one against accepting a poor solution as a good one. A large spread between λ_* and λ^* does not, however, necessarily indicate that λ^* is a poor solution. To be sure, the worst that can happen if λ^* is near-optimal and the spread between λ^* and λ_* is substantial is that one might continue to look for better solutions. More important, however, is the question of whether the technique does in fact produce near-optimal solutions.

Extensive examination of the Balinski-Mills method suggests that it is not well designed to handle the decreasing marginal cost functions generally postulated for the warehouse location problem. This can be illustrated by applying the method to the sample problem posed by Baumol and Wolfe. The Balinski-Mills solution calls for the use of all five warehouses at a cost of \$2,499 which is 22 percent above the optimal solution of \$2,047.

The lower bound (λ_*) to the solution of any problem provided by the Balinski-Mills method depends upon the quality of the solution used as a basis for calculation. The lower bound calculated on the basis of the Balinski-Mills solution to the Baumol-Wolfe problem is \$1,655, some 19 percent below the optimal solution. In contrast, the lower bound that would be calculated if the optimal warehouse network (namely, using only warehouse 3) had been identified is \$2,047, identical with the distribution cost of the warehouse network used as the basis for computing this value of the lower bound. Such an equality would insure that the optimal had indeed been attained. This pleasant state of affairs is not, however, guaranteed by the Balinski-Mills lower bound. In general the lower bound will be equal to the optimal solution only when each warehouse entering the solution is used to some predetermined capacity or alternatively, when the optimal solution consists of using only a single warehouse (that is, each warehouse in the optimal solution handles the volume a_j).

Mills, in personal communication, has stated that the Baumol-Wolfe problem violates one of the conditions set forth in the Balinski-Mills paper, namely, that the warehouse cost functions be "piecewise linear but not concave" ([2], p. 3).¹⁸ However, application of the method to such a

¹⁸The piecewise linear function illustrated by Balinski and Mills ([2], p. 4) is similar to the cost function in Figure 2 above. This function, with or without a fixed cost element, is concave (even strictly concave) over suitably specified regions. (For definitions of concavity, convexity, and the related extreme-point optimization properties, see Charnes and Cooper [7] pp. 284-87.)

problem does not seem inappropriate since the concave functions used in the Baumol-Wolfe problem can be closely approximated by piecewise linear functions containing the essential elements of the problem.

In comparing the Balinski-Mills and Baumol-Wolfe methods it should also be noted that the former permits the treatment of fixed costs whereas the latter appears to be unable to handle such cost elements. Comparing the two methods on a problem which does not contain fixed costs does not, therefore, fully test the Balinski-Mills method. However the difficulty with the method appears to stem from its inability to cope with functions which deviate significantly from simple linear functions. Thus, adding fixed costs might appear to aggravate the situation. For example, if there were fixed costs associated with operating warehouses in the Baumol-Wolfe sample problem, and if these costs were equal for each warehouse, the Balinski-Mills solution would be independent of the level of fixed costs. In contrast, we might logically expect to make use of fewer warehouses in a distribution system if the fixed costs per warehouse were increased. Consequently, the existence of fixed costs tends to increase the deviation of the Balinski-Mills solution from optimality.

Balinski [1] has applied this same average cost approximation technique to what he calls the fixed cost transportation problem. His results seem to show that the method, although not guaranteeing optimality, works fairly well in solving such problems. This problem is not the same, however, as the warehouse location problem. The essence of the difference between the two problems is that in the fixed cost transportation problem the function describing the fixed cost elements is a separable one. That is, the fixed cost associated with using any route (that is, any path from a source of supply to a point of demand) is independent of the other routes used. In the warehouse location problem, shipping and handling costs may take this form but, in addition, a fixed cost may be associated with the use of each warehouse. This cost is incurred if a warehouse is used, but is independent of the number of routes passing through that warehouse.

Simulation of a Distribution System

The use of simulation techniques in the modeling of warehouse networks has been proposed by Shycon and Maffei [25] as a means of avoiding the approximations in problem specification required by the techniques outlined above.

The value of a simulation approach to solving the warehouse location problem, much as with the use of any other technique, depends upon how well the model describes the essence of the system being studied and the time required in computation. As with the other approaches, some balance must be reached in terms of model detail and costs of computation. Once a model has been constructed containing the basic characteristics of the distribution system under study, it is used to evaluate the dis-

tribution costs associated with alternative sets of warehouse sites and customer ordering rates. The problem that remains is to devise an algorithm which can be used to generate at least one near-optimal warehouse system.

Shycon and Maffei do not outline a computational method for determining the sets of warehouse sites to be evaluated. Their approach requires that management or a consultant specify the warehouse systems (number of warehouses and their locations) to be evaluated. Maffei has stated that new warehouse systems can be generated from any given distribution network by moving individual warehouses fixed or random distances in random directions, but such a procedure does not appear to be an efficient method for searching for near-optimal warehouse location patterns.¹⁹

One other aspect of the Shycon-Maffei approach differs with respect to the other methods outlined in this appendix and the heuristic program itself. To save computer storage space they locate factories, warehouses, and customers in terms of longitude and latitude and subsequently approximate shipping costs between all of these points in terms of the computed air miles. The other methods outlined in this article would in general use actual costs, although a similar estimation technique could be used if actual data were not readily available. The use of air mile distances as approximations for shipping costs is an interesting approach to the problem and should be studied in more detail to determine the magnitude of error likely to result in cost data generated in this manner. If the error is within acceptable limits, it affords a means of reducing the time required in collecting and analyzing the cost data.

In its most detailed application, a simulation approach to the warehouse problem would permit the processing of actual customer orders through the alternative warehouse distribution systems being analyzed, and imputing costs to the various phases of the operation. This would be a desirable approach, particularly for seasonal products, since it would permit incorporation of the dynamic elements of warehousing and inventory costs. The Shycon-Maffei approach, like the other methods discussed in this paper, does not simulate the distribution system in that degree of detail but rather estimates the warehousing and inventory costs as functions of the total volume of each product routed through any given warehouse. (In the heuristic program, since the number of products considered can be increased with less than a linear effect on computation time, it is possible to describe customer demand in terms of discrete distribution of product mix and order size such that the program treats each order-size, product-mix combination as a separate product, except insofar as it influences inventory levels.)

In summary, the linear programming approaches to the warehouse location problem outlined in this section have as a desirable feature a systematic routine for generating alternative warehouse systems for evaluation.

¹⁹ Marketing Theory Seminar, University of Vermont, August, 1960.

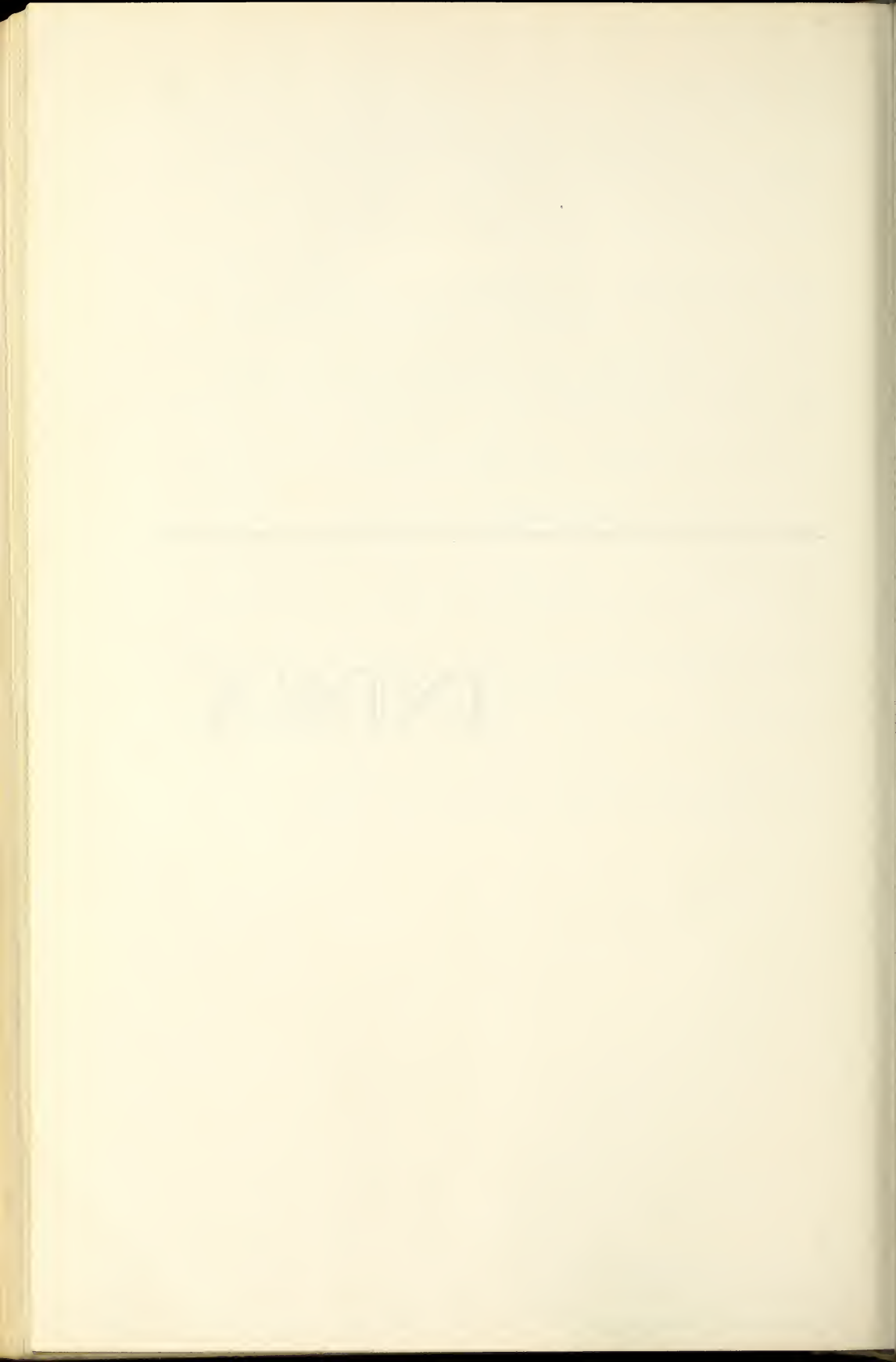
However, they are inferior to the simulation approach in that there is less flexibility available in the modeling of the distribution system. The computer program described in this paper combines most of the desirable features of simulation with a set of heuristics designed to generate near-optimal warehouse systems.

REFERENCES

1. BALINSKI, M. L. "Fixed-Cost Transportation Problems," *Naval Research Logistics Quarterly*, Vol. VIII (March, 1961), pp. 41-54.
2. BALINSKI, M. L., and MILLS, H. "A Warehouse Problem," prepared for: Veterans Administration; Mathematica, Princeton, New Jersey, April, 1960.
3. BAUMOL, W. J. *Economic Theory and Operations Analysis*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1961, pp. 410-13.
4. BAUMOL, W. J., and WOLFE, P. "A Warehouse Location Problem," *Operations Research*, Vol. VI (March-April, 1958), pp. 252-63.
5. BERNSTEIN, A.; ROBERTS, M. DE V.; ARBUCKLE, T.; and M. H. BELSKY. "A Chess-Playing Program for the IBM 704," *Proceedings of the 1958 Western Joint Computer Conference*, pp. 157-59.
6. CHARNES, A., and COOPER, W. W. "The Stepping Stone Method of Explaining Linear Programming Calculations in Transportation Problems," *Management Science*, Vol. I (October, 1954), pp. 49-69.
7. CHARNES, A., and COOPER, W. W. *Management Models and Industrial Applications of Linear Programming*. New York: John Wiley & Sons, Inc., 1961.
8. COURANT, R., and ROBBINS, H. *What Is Mathematics?* Oxford University Press, 1941.
9. CYERT, R. M.; DILL, W. R.; and MARCH, J. C. "The Role of Expectations in Business Decision Making," *Administrative Science Quarterly*, December, 1958.
10. DANTZIG, G. B. "Application of the Simplex Method to a Transportation Problem," T. C. Koopmans (ed.), *Activity Analysis of Production and Allocation* (Cowles Commission Monograph No. 13). New York: John Wiley & Sons, Inc., 1951, chap. xxiii.
11. FORD, L. R., and FULKERSON, D. R. "Solving the Transportation Problem," *Management Science*, Vol. III (1956), pp. 24-32.
12. GELERNTER, M., and ROCHESTER, N. "Intelligent Behavior in Problem-Solving Machines," *IBM Journal of Research and Development*, Vol. II (October, 1958), pp. 336-45.
13. GOMORY, R. E. "An Algorithm for Integer Solutions to Linear Programs," Princeton-IBM Mathematics Research Project, Technical Report No. 1, Princeton University.
14. GOODWIN, G. H. "Digital Computers Tap Out Designs for Large Motors Fast," *Power* (April, 1958).
15. HILLER, L. A., and ISAACSON, L. M. *Experimental Music*. New York: McGraw-Hill Book Co., Inc., 1959.
16. HOLT, C. C.; MODIGLIANI, F.; MUTH, J. F.; and SIMON, H. A. *Planning Production, Inventories, and Work Force*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1960.

17. KEMENY, J. G., and THOMPSON, G. L. "The Modified Fictitious Play Method," *Dartmouth Mathematics Project Report*, 1958.
18. KISTER, J.; STEIN, P.; ULAM, S.; WALDEN, W.; and WELLS, M. "Experiments in Chess," *Journal of the Association for Computing Machinery*, Vol. IV (April, 1957), pp. 174-77.
19. MARCH, J. G., and SIMON, H. A. *Organizations*. New York: John Wiley & Sons, Inc., 1958.
20. NEWELL, A.; SHAW, J. C.; and SIMON, H. A. "Chess Playing Programs and the Problem of Complexity," *IBM Journal of Research and Development*, Vol. II (October, 1958), pp. 320-35.
21. NEWELL, A.; SHAW, J. C.; and SIMON, H. A. "The Processes of Creative Thinking," the RAND Corporation Paper, P-1320, August, 1958.
22. NEWELL, A., and SIMON, H. A. "The Logic Theory Machine," *IRE Transactions on Information Theory*, IT-2 (September, 1956), pp. 61-79.
23. *Rand-McNally—Cosmopolitan World Atlas*. Chicago: Rand-McNally & Co., 1951.
24. SAMUEL, A. L. "Some Studies in Machine Learning Using the Game of Checkers," *IBM Journal of Research and Development*, Vol. III (July, 1959), pp. 210-29.
25. SHYCON, H. N., and MAFFEL, R. B. "Simulation—Tool for Better Distribution," *Harvard Business Review*, November-December, 1960, pp. 65-75.
26. SIMON, H. A. *Models of Man*, New York: John Wiley & Sons, Inc., 1957.
27. SIMON, H. A. "Modeling Human Mental Processes," *Proceedings of the 1961 Western Joint Computer Conference*.
28. TONGE, F. M. *A Heuristic Program for an Assembly Line Balancing Problem*. New York: Prentice-Hall, Inc., 1961.
29. TONGE, F. M. "The Use of Heuristic Programming in Management Science," *Management Science*, Vol. VII (April, 1961), pp. 231-37.
30. *The World Almanac*. New York: New York World-Telegram and the Sun, 1960.

INDEX



Index

A

Acton, F. S., 272
 Advance orders, 506 ff
 Advertising, 134
 budgets, 192-93
 effect, measurement of, 190-203
 effectiveness, 167-76
 expenditure, 274 ff
 mood and illusion, 167-76
 point of purchase, 274 ff
 remembrance, 167-76
 sales claim, 167-76
 television, 166-76
 themes, effects of, 204-19
 Advertising Research Foundation, 427, 428, 429, 436, 438
 Alderson, Wroe, 129, 132 fn, 140, 143
 Algorithm, 115 ff, 524
 American Documentation Institute, 429, 430
 American Medical Association, 246
 Analysis of variance model, 206, 214
 Anderson, R. L., 10 fn, 60 fn, 80 fn, 85 fn, 272, 460 fn
 Anderson, T. W., 375 fn
 Anscombe, Frank J., 347 fn, 356
 Arrays, 60 ff
 advantages of, 64
 Arrow, Kenneth J., 5 fn
 Association, 147 ff
 linear, 72 ff, 101
 spurious causal, 56
 statistical, 56 ff
 Attributes table, 62
 Attwood Consumer Panel, 355, 361
 Audience Research, 225
 Audit Bureau of Circulation, 4, 427
 Autocorrelation, 87

B

Baker, G. A., 409
 Baker, Henry G., 306
 Balderston, F. E., 122
 Balinski, M. L., 537, 538, 540-43
 Bancroft, T. A., 60 fn, 80 fn, 85 fn, 272, 460 fn
 Banks, Seymour, 9, 270, 273
 Baron, Margaret, 410, 416, 422

Bartlett, M. S., 182, 272, 459 fn
 Baumol, W. J., 525 fn, 537-40, 542, 543
 Bayesian decision theory, 17 fn
 Before-and-after design, 220-27
 Behavior
 competitive, 58
 consumer, 129-45
 habitual purchasing, 374 ff
 rational consumer, 5, 139
 Behavior research, 130-31
 Behavioral relation, 113
 Berelson, Bernard R., 474 fn
 Bernard, Claude, 237 fn
 Bias, 59, 66, 85
 Biased estimate, 66, 82, 88
 Biased estimator, 83, 85
 Blair, G. W. Scott, 410, 416, 426
 Blake, Robert R., 318 fn
 Bloom, B. S., 317 fn
 Bonnick, Gladstone, 333 fn
 Borden, N. H., 279
 Brand
 acceptability, 270
 choice, 372-89
 image, 326-31
 loyalty, 331-33
 preference, 156 ff, 273 ff
 purchase sequences, 394 ff
 runs, 375 ff
 share, 269 ff
 shifting model, 390 ff
 Britt, S. H., 437
 Brown, D. A., 361
 Brown, George, 9, 372
 Brunk, Max E., 51 fn
 Bureau of Applied Social Research, 326
 Bush, Robert R., 375, 390, 392, 393
 Bush-Mosteller stochastic learning
 model, 390 ff
 Business game, 118 ff
Business Periodicals Index, 24

C

Carry-over effects, 50, 195-98, 205 ff
 Causal
 analysis, 82
 designs, 22, 26
 hypothesis, 27
 link, 56

- Causal—*Cont.*
 relationships, 22 ff
 research, 26
 studies, 26–28
- Causality, 57, 113
- Cell by cell balancing, 88
 desirability of, 66–69
- Centroid matrix, 432
- Centroid method of factor analysis, 418
- Chance variation, 7
- Charnes, A., 539 fn
- Chicago Tribune*, 21, 273
- Chicago Tribune* consumer panel, 373, 394
- Chi-square, 149, 188
- Churchill, Harold E., 307
- Cochran, William G., 36 fn, 43 fn, 150 fn, 204, 211, 213
- Cochran's criterion, 150
- Coefficient of elasticity, 163 fn
- Coefficients of determination, 432
- Coleman, James S., 258 fn, 259 fn, 375 fn
- Collinearity, 444
- Collins, M. E., 46
- Communality of the test, 414
- Comparisons, 11, 45
- Competitive behavior, 58
- Competitive Preference technique, 166
- Competitive responses, 12
- Computer simulation, 465 ff
- Concomitant variation, 28–30, 35
- Confidence interval, 59, 160 fn
- Consumer
 behavior, 129–45
 brand shifting, 390–403
 irrationality, 140
 point of view, 16
 purchases, 355–71
 rational behavior, 5, 139
- Consumer panel, 47
 Attwood, 355, 361
Chicago Tribune, 373, 394
 National, 345, 353
 Thompson Company, 338
- Contagion, spurious, 378 ff
- Contamination of control units, 51
- Control group, 33 ff
- Controlled environment, 154 ff
- Controlled experiment, 192
- Cook, Stuart, 21 fn, 23 fn, 45 fn
- Cooper, W. W., 539 fn
- Correlation, 56, 60, 104, 151, 429 ff
 analysis, 69–78
 coefficients, 69 ff, 98, 411 ff, 432 ff
 matrix of, 101
 multiple, 77, 78, 274 ff
 partial, 75 ff, 416
 simple, 70 ff
 matrix, 74, 93, 412, 429 ff, 444
 multiple, 77–78, 221 ff
- Correlation—*Cont.*
 partial, 74–77, 416
 simple, 70–74
 spurious, 56
- Courant, R., 524
- Covariance, 70
 analysis of, 42, 210 ff
 model, 214
- Covariation, 70, 72
- Cox, D. R., 36 fn, 43 fn
- Cox, Gertrude M., 36 fn, 43 fn
- Cronbach, L. J., 416
- Cross-classification, 11, 61–69, 104
 table, 60 ff
- Cross elasticities, 293 ff
- Cross tabulation, 233
- Cross-over design, 205
- Cross-sectional studies, 34, 46–47
- Cumulative curve, 240 ff
- Cunningham, Ross M., 372
- Cushen, Walter E., 118 fn
- Cyert, Richard M., 524 fn
- D**
- Davidson, Donald, 154 fn
- Day, Ralph L., 111, 119 fn
- Decision, 3
 function, 113
 making, 3, 22
 units, 481 ff
 theory
 Bayesian, 17 fn
 statistical, 40
- Degrees of freedom, 149 ff, 208 ff
- Demand
 curves, 158 ff
 functions, 282 ff
 model, 269 ff
 schedules, 158 ff
- Demographic variables, 318 ff
- Dependent variables, 29, 35, 56 ff, 272 ff
- Dernburg, Thomas F., 441 ff
- Descriptive research, 26
- Descriptive studies, 25–26
- Designs
 before-and-after, 220–27
 causal, 22, 26
 cross-over, 205
 descriptive, 21, 26
 experimental, 35, 57 ff, 205 ff, 428
 exploratory, 21
 observational, 57
 switch back, 205
 switch-over, 193 ff, 205
 types of experimental; *see* Experimental design, types of
- Deutsch, Morton, 21 fn, 23 fn, 45 fn, 46
- Dill, William R., 524 fn
- Discriminant
 analysis, 61, 74, 95–100, 104

Discriminant—*Cont.*
 analysis—*Cont.*
 multiple, 95–100
 boundary line, 96 ff
 intercept of, 99
 slope of, 99
 function, 96 ff
 linear function, 304, 308, 311 ff
 Dispersion, 59
Dissertation Abstracts, 24
 Distribution, 147 ff
 negative binomial, 356 ff
 Poisson, 364
 of runs approach, 385–89
 Dixon, Wilfred J., 70 fn

E

Economic Survey of Europe, 282
 Edwards, Allen L., 309 fn, 330 fn, 337, 338
 Edwards Personal Preference Schedule,
 49, 309, 330
 Effects, carry-over, 50, 195–98, 205 ff
 Efficiency in an estimate, 85
 loss of, 67
 Eisenhart, Churchill, 206
 Elasticity
 coefficient of, 163 fn
 cross, 293 ff
 income, 282 ff
 price, 159 ff, 293 ff
 Eldersveld, Samuel, 469
 Empirical data, 59 ff
Encyclopedia of American Associations, 24
 Enumeration table, 62
 Environment, controlled, 154 ff
 Environmental variables, 114
 Error variance, 86
 Errors in equations, 83, 87
 Errors in variable models, 83, 87
 Estes, William K., 390
 Estimate
 biased, 66, 82, 88
 statistical, 59 ff
 unbiased, 60, 84, 85, 90, 100
 Estimating demand, 153–65
 Estimator
 biased, 83, 85
 efficient, 60
 Evans, D. A., 356, 357
 Evans, Franklin B., 49, 305 fn
 Exogenous variables, 114
 Experiment, controlled, 192
 Experimental, 6, 7
 design, 35, 57 ff, 205 ff, 428
 types of, 36–40
 after-only design, 36–37
 before-after design with one control group, 37–38
 before-after design with two control groups, 38–40

Experimental—*Cont.*
 group, 33 ff
 measurement, expense of, 50
 method, 34
 variable, 36 ff
 Experimentation, 10, 33, 35–44, 270
 limitations of, 49–52
 Explanatory
 designs, 21
 studies, 23–25, 26
 variables, 56 ff, 77, 105, 296, 440 ff
 Extraneous
 factors, 57
 variables, 57
 Eysenck, H. J., 411

F

F test, 211, 213
 Factor
 analysis, 7, 61, 74, 100–104, 409–26
 loadings, 102–4, 418, 432 ff, 444 ff
 loadings matrix, 446 ff
 matrix, 412, 413
 scores, 444 ff
 derivation of, 458–60
 Factor analytic study, 440–60
 Factorial design, 44, 410
 Family interactive patterns, 143–45
 Family panel, 131
 Federer, Walter T., 36 fn, 43 fn, 204, 207
 Feedback, 112, 114
 Feller, William, 375 fn, 380 fn
 Ferber, Robert, 12 fn
 Ferguson, L. W., 437
 First repeat ratio, 346
 Fisher, R. A., 308 fn, 311 fn, 313 fn, 338
 Fisher's *z* transformation, 436
 Flesch, R., 429
 Flow-of-funds system, 497
 Food consumption trends, 281–302
 Food prices, influence of changes in,
 291–93
 Forecast error, 79
 Forecasting, 79 ff
 Forrester, Jay W., 110 fn, 114
 Frank, Ronald E., 401, 402
 Friedman, Milton, 5 fn, 6 fn
 Furst Survey Research Center, 470

G

Gallup, George, 47, 427, 465 fn, 470
 Garrett, Henry E., 338
 Girshick, M. A., 288 fn
 Gleser, G. C., 416
 Gomory, R. E., 537
 Goodman, Leo A., 337 fn
 Goodness-of-fit test, 149
 Grant, Eugene L., 14 fn

H

- Haavelmo, T., 288 fn
 Habitual purchasing behavior, 374 ff
 Harmon, Harry, 100 fn, 101 fn, 459 fn
 Harper, Roland, 410, 416, 418
 Hawkins, Edward R., 153 fn
 Hayakawa, S. I., 307
 Henderson, Peter L., 204
 Heron, David, 337
 Herrmann, Cybil C., 154 fn
 Heterogeneity, 149 ff
 Heteroscedasticity, 86
 Heuristic(s), 106, 115, 524 ff
 approach, 115 ff, 524
 computer program, 523 ff
 program for locating warehouses, 525 ff
 bump and shift routine, 525 ff
 main program, 525 ff
 programming, 114-18, 524 ff
 techniques, 524
 Hodges, Joseph L., 338
 Hoel, Paul G., 95 fn
 Hoggatt, A. C., 122
 Holt, C. C., 537 fn
 Holzinger, Karl J., 446
 Homoscedastic, 86, 459 fn
 Hood, William C., 61 fn, 82 fn
 Hotelling, H., 424
 Hotelling-Jur  en relation, 297
 Howard, John A., 21 fn, 41 fn, 49 fn
 Hughes, H. M., 38 fn
 Hypothesis
 causal, 27
 meaningful, 23
 substantive, 4, 5

I

- Identification of problems, 13
 Identities, 483 ff
 Income elasticity, 282 ff
 Independent variables, 29, 56 ff, 113, 114, 272 ff, 444 ff
 Index of
 homogeneity, 262
 simultaneity, 259, 260
 Industry trade publications, 24
 Information, concept of, 58-60
 Information feedback system, 114
 Interaction, 12, 36, 38, 39, 44 fn, 68, 69, 134, 183, 395
 Interactive models, 106-23
 Intercorrelation, 310, 319 ff, 429, 430
 Intuition, 4, 14, 15

J

- Jahoda, Marie, 21 fn, 23 fn, 45 fn
 Janowitz, Morris, 469
 Jessen, Raymond J., 50, 204
 Jeuck, John E., 3 fn, 10, 13

Johnson, Samuel, 436

Journal of Advertising Research, 204
 Jur  en, Lars, 281 fn, 290 fn, 294 fn, 295, 297 fn

K

- Kaiser, Henry F., 446
 Katona, George, 140 fn
 Katz, Elihu, 48, 258 fn, 259
 Kellogg, Lester S., 59 fn, 85 fn
 Kendall, M. G., 410
 Klein, Laurence R., 84 fn, 92 fn
 Knight, Frank H., 4 fn
 Knutsen, Mack, 338 fn
 Koopmans, Tjalling C., 61 fn, 82 fn
 Koponen, Arthur, 336 fn, 338 fn
 Kreinen, Mordechai E., 305 fn
 Kruskal, William H., 337 fn
 Kuehn, Alfred A., 111, 351, 373, 378 fn, 393 fn, 401 fn

L

- Laboratory sensory-difference trial, 151
 Laboratory technique, 168
 Lane, Robert, 469
 Lansing, John B., 305 fn
 Large sample interviews, 130
 Lasswell, Harold, 469
 Latin square, 41, 42, 195, 205, 211
 Law of Social Change, 257 fn
 Lazarsfeld, Paul F., 30, 48, 469, 474 fn
 Least significant difference test, 211, 213
 Least squares analysis, 79 ff
 Leavitt, Harold J., 309 fn
 Leontief input-output matrix, 489
 Levenson, Bernard, 326 fn
 Linear
 association, 72 ff, 101
 combination, 77 ff, 95, 304
 dependence, 69, 74
 discriminant function, 304, 308, 311 ff
 regression, 273
 Linearity, 64, 88
 Lipstein, Benjamin, 354 fn
 Logical and mathematical tools, 4, 6
 Loric, James H., 9 fn, 12 fn
 Lucas, D. B., 207, 427, 437
 Lymberopoulos, P. John, 119 fn

M

- Maffei, Richard B., 112, 526, 543, 544
 Management game, 118
 Management science, 115
 March, James G., 524 fn
 Markdown pricing, 513 ff
 Market Research Corporation of America, 47
 Market research methods, 304-5
 motivation research, 130, 131, 305
 traditional research, 304

- Market, share of, 269 ff
 - Marketing management, 3, 22
 - Marketing research, 15, 16
 - Markov, 82
 - Markov chain analyses, 395
 - Mark-up goal, 505 ff
 - Martineau, Pierre, 306
 - Masek, John, 305 fn
 - Massey, Frank J., Jr., 70 fn
 - Massy, William F., 30 fn, 32 fn, 43 fn, 441 fn, 443 fn, 444 fn
 - Matching, 34 ff
 - technique, 221
 - Matrix of correlation coefficients, 101
 - Maximum likelihood, 356
 - procedures, 61 fn
 - McPhee, William, 468, 474 fn
 - Measuring advertising effect, 190-203
 - Measuring technique, 168-71
 - Menzel, Herbert, 258 fn, 259 fn
 - Meredith, G. P., 425
 - Mickey, M. R., 200 fn
 - Mills, H., 537, 538, 540-43
 - Misclassification, 99
 - Missing-observation problem, 52
 - Model(s), 57 ff, 106 ff
 - brand shifting, 390 ff
 - building, 106 ff
 - classification of, 108-10
 - concept of, 106-8
 - descriptive, 110
 - deterministic, 109
 - dynamic, 108 ff
 - empirical, 109
 - interactive, 106-23
 - market demand, 269-80
 - microanalytic, 481-501
 - normative, 109
 - recursive, 491 ff
 - regression, 213
 - simulation, 502 ff
 - static, 108, 112
 - stochastic, 109, 111
 - theoretical, 109
 - Modigliani, F., 537 fn
 - Monte Carlo approach, 381-85
 - Monthly Catalog of United States Government Publications*, 24
 - Morgan, James N., 305 fn
 - Mosteller, C. Frederick, 375, 390, 392, 393
 - Motivation research, 130, 131
 - large sample interviews, 130
 - small sample studies, 130
 - Mouton, Jane S., 318 fn
 - Mueller, Eva, 140 fn
 - Multicollinearity, 93, 101, 444
 - Multiple
 - correlation, 77-78, 221 ff
 - covariance analysis, 213 ff
 - discriminant analysis, 95-100
 - Multiple—*Cont.*
 - discriminant boundary line, 96 ff
 - linear regression, 273
 - regression, 11
 - Munn, Henry L., 326
 - Murray, Henry A., 309
 - Muth, John F., 537 fn
- N**
- National Consumer Panel, 345, 353
 - National Institute for Research in Dairy-
ing, 410
 - National Opinion Research Center, 52
 - Negative binomial distribution, 356 ff
 - Newell, A., 115, 524
 - Nielson, Arthur C., Jr., 167, 345 fn
 - Nielson Company, A. C., 47, 273
 - Nonbrand runs, 376 ff
 - Nonexperimental, 6
 - Normal deviate, 385 ff, 399
 - Normal pricing, 511 ff
- O**
- Observational, 6
 - design, 57
 - studies, 11, 33-35, 44-49
 - techniques, 11
 - One-store shopper, 141-42
 - Open-to-buy, 509 ff
 - Operating characteristic, 112, 113, 483 ff
 - Operational gaming, 106, 118-20
 - Operations research, 115, 135, 524
 - Ordering decisions, 502 ff
 - Ostle, Bernard, 213, 409
 - Ownership index, 450
 - profile, 451 ff
- P**
- Paired comparison, 151, 186, 309
 - Panel studies, 46, 47, 48
 - Parameters, partial regression, 92
 - Parsimony of variables, principle of, 58
 - Partial correlation, 416
 - Pearl-Reed growth curve, 443
 - Pearson, Karl, 337
 - Penetration, 346
 - prediction, 347-50
 - Pessimier, Edgar A., 154 fn
 - Pike, Eugene W., 8 fn
 - Placebo effect, 52
 - Point of purchase advertising, 274 ff
 - Point in time analysis, 34
 - Poisson distribution, 364
 - Politz, Alfred, 11, 46, 48
 - Pooled data, 68
 - Postpurchase matters, 132
 - Prediction of sales, 350-54
 - Preference, 146-52
 - measurement, 9

Preference—*Cont.*

ratings, 272 ff

tests, 179–89

Prepurchase matters, 132

Price, 274 ff

Price elasticity, 159 ff, 293 ff

Pricing, 134

decisions, 502 ff

markdown, 513 ff

normal, 511 ff

sale, 512 ff

Probability function, 483 ff

Problem, 3

identification of, 13

Product line, 134 ff

decisions, 135–39

assortment, width versus depth of,
134, 135, 136–37

exclusiveness of, 138–39

number of items, 134, 135

price line, width of, 134, 135, 137–38

Projective tests, 309

Promotional effort, 274 ff

Proximity theorem, 88

Purchase operator, 391 ff

Psychological

equivalence, 154, 155

factors, 49

scaling, 7

testing, 305–6, 308–10

variables, 309 ff

Q

Quality control, 14

R

Random, 7

selection, 272

Randomization, 7, 10, 34, 36, 37, 41

Randomized experimentation, 10

Randomly distributed, 85

Range of overlap, 99

Rao, C. R., 338 fn

Rational consumer behavior, 5, 139

Readership recognition surveys, 427 ff

Recursive model, 491 ff

Refusal rate, 48

Regression, 60, 104

analysis, 74, 78–95, 270 ff

multiple, 11, 92–95, 214

univariate, 79–92

assumptions, 85–90

coefficients, 83 ff, 274 ff

multiple, 94

partial, 83, 98, 100

simple, 83

equation, 221

error term, 83, 85

method, weighted, 92

model, 213

Regression—*Cont.*

residuals, 90–92

slope parameter, 79 ff

Re-interview bias, 48

Reisman, David, 335

Rejection operator, 391 ff

Reorders, 509 ff

Repeat purchase probability, 374 ff

Repeat ratios, 350 ff

Repeated surveys, 47

Research, 4, 13–17

analysis, 233–38

design, 21–55, 220

Response

coefficient, 82

competitive, 12

function, 112–14

long term, 49

sales, 197 ff

short term, 49, 50

Return purchase probability, 374 ff

Ridenour, Louis N., 257 fn

Robbins, H., 524

Roberts, Harry V., 9 fn, 11 fn, 12 fn,

13 fn, 16 fn, 21 fn, 41 fn, 49 fn, 323 fn,

385 fn

Roper, Elmo, 47, 465 fn

Roper Public Opinion Research Center,

465, 469

Runs, 375 ff

brand, 375 ff

distribution of, 385–89

nonbrand, 376 ff

S

Sale pricing, 512 ff

Sales

analysis, 11

goal, 503 ff

prediction of, 350–54

response, 197 ff

tests, 179–89

variability of, 50

Sample moments, 98 ff

Scatter diagram, 60 ff, 281

Schlaifer, Robert, 17 fn, 28 fn, 60 fn

Schweiger, Irving, 12 fn

Selling facilities, 134

Sellitz, Claire, 21 fn, 23 fn, 45 fn

Sessions, Robert E., 129, 132 fn, 140, 143

Share of market, 269 ff

Shaw, J. C., 115, 524

Shewhart, Walter A., 14 fn

Shycon, Harvey M., 112, 526, 543, 544

Siegal, Sidney, 154 fn

Significance of the difference, 182 ff

Simon, Herbert A., 5 fn, 27 fn, 115, 524,

537 fn

Simulation(s), 106, 108, 110–14, 472–80,

484 ff

- Simulation(s)—*Cont.*
 computer, 465 ff
 model, 502 ff
 shopping trip, 154 ff
 Simulmatics Corporation, 469 ff
 Slope parameter, regression, 79 ff
 Slutsky-Schultz relation, 299 fn
 Small Business Administration, 24
 Small sample motivation research studies, 130
 Smith, Babington, 410
 Smith, John H., 59 fn, 85 fn
 Snedecor, George W., 77 fn, 182
 Social Research, Inc., 21
 Sociogram, 251
 Spurious causal association, 56
 Spurious contagion, 378 ff
 Spurious correlation, 56
 Spurr, William A., 59 fn, 85 fn
 Square-root transformation, 182
 Stable correlates, 222 ff
 development of, 222-24
Standard Advertising Register, 24
 Standard error, 184, 185, 360
 Standard and Poor's *Industry Surveys*, 24
 Standard Rate and Data Service, 24
 Standardization, 70
 Standardized averages, 11
 Star, S. A., 38 fn
 Starch, Daniel, 427
 Statistical
 allowance, 11
 association, 56 ff
 decision theory, 40
 estimate, 59 ff
 independence, 87
 inference, 6, 41, 59
 techniques, 56-105
 Stein, M. I., 317 fn
 Stern, G. G., 317 fn
 Stewart, John B., 154 fn
 Stochastic, 483
 learning model, 390 ff
 Stock display, 274 ff
 Stone, R., 288 fn, 295
 Store clusters, 142-43
 Store coverage, 274 ff
 Straight-run tabulations, 233
 Structural analysis, 82 ff, 87
 Suits, Daniel B., 319 fn
 Sum of squares, 201, 207 ff
 Suppes, Patrick, 154 fn
 Survey Research Center of the University of Michigan, 140, 141
 Switch-back design, 205
 Switch-over design, 193 ff, 205
 Systematic error, 7
 Television advertising, 166-76
 Test group, 33 ff
 Tests of significance, 272
 Thematic Apperception Tests, 131, 143, 309
 Thiel, Henry, 61 fn
 Thomas, George B., Jr., 102 fn
 Thompson, J. Walter, advertising agency, 436
 Thompson Company, J. Walter, Consumer Panel, 338
 Thurstone, L. L., 418
 Thurstone's centroid method, 432
 Time order of variables, 31, 35, 46
 Time rate of decay of purchase probability, 396
 Time series analysis, 34, 46
 Tintner, Gerhard, 82 fn, 87 fn, 95 fn, 102 fn
 Tischer, R. G., 409
 Tobin, J., 288 fn
 Tonge, F. M., 524
 Transformation, 357, 443
 of variables, 64, 90
 Transition, 108, 109
 Treatments, 41 ff, 193, 205 ff
 Tukey, John, 469
 TV's Law of Extremes, 167
 Two stage least squares method, 61 fn
- U**
- Unbiased
 estimates, 60, 84, 85, 95, 100
 forecast, 82
 Uncontrolled variables, 37 ff, 50, 57
 Uncontrolled variation, 50
 United Nations, 38
 United States Brewers Foundation, 21
 University research projects, 24
- V**
- Valavanis, Stefan, 85 fn
 Variability of sales, 50
 Variables, 23, 482 ff
 demographic, 318 ff
 dependent, 29, 35, 56 ff, 272 ff
 environmental, 114
 exogenous, 114
 experimental, 36 ff
 explanatory, 56 ff, 77, 105, 296, 440 ff
 extraneous, 57
 independent, 29, 56 ff, 113, 114, 272 ff, 444 ff
 input, 482 ff
 output, 482 ff
 psychological, 309 ff
 socioeconomic, 309 ff
 status, 482 ff
 time order of, 31, 35, 46
 uncontrolled, 37 ff, 50, 57
- T**
- Tasting experiment, 146-52

Variance, 59 ff
 analysis of, 182 ff, 207 ff
 Vincent, D. F., 411

W

Wallace, David, 307
 Wallis, W. Allen, 11 fn, 16 fn, 323 fn,
 385 fn
 Warehouse location problem, 523-46
 Washington State Apple Commission,
 204
 Washington State University, 155
 Wearmouth, W. G., 422, 426
 Weighted regression method, 92
 Weiss, Doyle L., 120 fn

Wherry-Doolittle formula, 222
 Wilkins, L. T., 221
 Wold, Herman, 88, 281 fn, 290 fn, 294 fn,
 295, 297 fn
 Wolfe, P., 525 fn, 537-40, 542, 543
 Womer, Stanley, 346 fn
Wood Chips, 204, 236
 Woolf, James D., 436, 437

Y

Yule, G. Udny, 337

Z

Zeisel, Hans, 21 fn, 30, 47 fn

This book has been set in 10 on 12 and 9 on 10 Janson. Part numbers are 18 and 36 point Caslon Open Face. Part titles are 24 point Spartan Heavy. Article titles are 18 point Spartan Medium. The size of the type page is 27 x 47 picas.







No.	Returned	Due	Re' - ne'
-----	----------	-----	-----------

Date Due

[illegible]

658,83

F8283

BUSINESS

ADM.

Quantitative techniques in mar main
658.83F828q



3 1262 03310 2913

800-7

